About time: Temporality in interaction

A consolidation of research on the field of system delays in interactive systems and an experiment studying the effects of constant sub-second delays on human operators

Martin Krampell
Abstract

Ever since the inception of the modern computer, researchers and designers alike have been interested in the effects of system delays on users. The current study was conducted in order to examine the most central issues to the field of temporality in interaction, and presents a consolidation of a selection of publications on the subject. A distinction between two types of interactive systems, discretionary and continuous, is proposed in order to situate previous studies by the system being studied. The type of control being exerted by users differs on a fundamental level between the two types, hence affecting the effects of delays. Furthermore, an experiment was conducted to examine the effects of constant, sub-second system delays in discretionary tasks using a digitalised version of the Trail Making Test (FR-TMT, Summala et al., 2008).

The experiment yielded but one significant result in form of an improvement in user response time as delays were increased. The other results showed no significant positive or negative effect of increased delays. These results are indicative that the chosen delays do not have any detrimental effects on users, in accordance with the presently coined ‘theory of task interruption’. This theory considers delays as either interruptive or non-interruptive and maintains that only delays that disrupt user work-flow are to be removed from interactive systems. The current study gives reason to why some delays can be positive to user interaction, or in themselves be informative of system status, and be an integral part of a feedback structure.

Further research is needed before all aspects of system delays are fully understood. New ways of looking at delays and using them in system design, like predictability and predictivity, are becoming more prevalent, and may become the focus of research and temporal design in the near future.
Acknowledgements

First and foremost, I would like to express my deepest gratitude to my father, a true stoic and paragon of virtue, for having the patience to listen through my endless ramblings, and for helping me structure my work, internally or otherwise. My family, friends and colleagues deserve their own thanks, for support, and good humour throughout my work.

I wish to extend my sincere thanks to my supervisor, Rachel Ellis, for her continuous dedication and support throughout my project. An honourable mention goes to both Arne Jönsson and Magnus Bång, for their help and consultation along the way.

I also wish to thank inUse Experience AB for their support and collaboration in making this thesis a reality. I am grateful to Kjell Persson, for believing in me and providing me with this fantastic opportunity to investigate a subject I personally find very intriguing. I would also like to thank Fredrik Andersson, for giving me a glimpse into the inner workings of interaction design and engineering in the UX industry. Furthermore, I thank Rebecca Hallqvist for providing external review of a first draft.

I am indebted to The Swedish National Road and Transport Research Institute (VTI), and especially Björn Peters at VTI, for taking an interest in my project, and later providing me with equipment and a venue for conducting my experiment. I would also like to thank Carina Fors and Jan Andersson for their very helpful assistance in this endeavour.

I thank Heikki Summala for permission to use their digital version of the Trail Making Test, and for insightful comments regarding use of the test. I also wish to thank Torbjörn Kvist, for his most helpful assistance in re-programming the test to fit my needs.

I would also like to thank Rasmus Bååth, and that thank is two-fold. Firstly, for introducing me to the exciting field of cognitive science, and secondly, for casually proposing the notion that eventually grew to become this very thesis.

Finally, I wish to thank everyone that participated in my experiment, and so provided me with solid data.
Preface

This bachelor’s thesis was realised in collaboration with inUse Experience AB, and is the final assignment of three years of study at Linköping University, on the Cognitive Science program. inUse is a digital design agency, striving to create a better user experience across many types of systems and solutions, and were interested to have the subject of temporality in interactive systems investigated.

It was therefore not only in my interest to deliver a scientifically sound thesis, but to produce a coherent consolidation of the existing empirical evidence, in the hope of yielding a useful and constructive report on the subject. It was also my wish to write this thesis with the perspective and mindset of a cognitive scientist (if ever such a person really exists) and perhaps to contribute something new and unique to the field.

Much has happened during the five months it has taken to complete this project. Personally, this has been a fantastic learning experience, from which I have acquired many skills, be they explicit or implicit. This thesis also counts as my first major scientific text written in English, which has proven to be very exciting, but also quite challenging. The most eye-opening aspect has been the sheer magnitude of the project, and that I was able to see it through in the end. Believe me, it seemed insurmountable at the beginning, and still remains quite surreal.

Another interesting insight I can share is the – initially obscured, but now apparent – complex nature of the studied subject. What I initially thought would be a rather simple relationship between time and interaction, proved to be much more than that. Perhaps it was naïve of me to even believe that any problem involving humans and technology would be simple. There is a great lesson to be had from this; in that nothing is as straightforward as it might first seem. I believe these insights are indications that the writing of this thesis has succeeded in conveying at least one of its scholarly purposes.

All in all, my greatest aspiration is that this work can contribute to the existing research, and help expand important issues in the field of cognitive science, as well as in the industry. It is, after all, a testament to three years of study, and hopefully, the beginning of an interesting career. There is yet much to learn, and I look forward to what the future might hold.

Martin Krampell
Linköping, June 25, 2014
# Contents

1 Introduction .................................................. 1  
  1.1 Purpose ................................................. 2  
  1.2 Research questions ...................................... 3  
  1.3 Research approach ...................................... 3  
  1.4 Delimitations ........................................... 4  

2 Background .................................................. 5  
  2.1 The foundation of this study .............................. 5  
    2.1.1 The human operator ................................. 5  
    2.1.2 Two types of interactive systems .................. 6  
    2.1.3 Defining temporality ................................ 8  
  2.2 Review of literature ..................................... 9  
    2.2.1 General theoretical framework ..................... 9  
    2.2.2 Discretionary systems .............................. 16  
    2.2.3 Continuous systems .................................. 18  
    2.2.4 Summary of reviewed literature .................... 22  
  2.3 Outlining the experiment ................................ 23  

3 Method ...................................................... 24  
  3.1 Participants ............................................ 24  
  3.2 Apparatus ............................................... 24  
  3.3 Procedure .............................................. 25  

4 Results ...................................................... 27  
  4.1 Response time .......................................... 27  
    4.1.1 Subtask A Fixed ..................................... 27  
    4.1.2 Subtask A Random ................................... 28  
    4.1.3 Subtask B Fixed ..................................... 29  
    4.1.4 Subtask B Random ................................... 29  
  4.2 Error-rate ............................................... 30  

5 Discussion .................................................. 31  
  5.1 Discussion of experiment ............................... 31  
    5.1.1 Ecological validity .................................. 32  
    5.1.2 Limitations ........................................... 33  
  5.2 General discussion ...................................... 34  
    5.2.1 Theoretical implications ............................ 34  
    5.2.2 Practical implications ............................... 36  
  5.3 Are system delays soon a thing of the past? .......... 38  

6 Conclusions ................................................. 40  

7 Future research ............................................. 41  

References .................................................... 42
1 Introduction

In our ever-evolving digital age, computers and interactive media have become the norm, rather than the exception it might have been only a few decades ago. With this new spring of technology, comes a range of unprecedented and previously unforeseeable new issues. This is the flip side of technology, that in solving one problem, one inevitably creates a new one—granted, often less severe. These problems are often related to the interaction with technology (as in bad design), rather than there being anything wrong with the solution itself.

These fundamental aspects of modern computer science are at the core of the scientific fields of human factors, as well as human-computer interaction (HCI). These fields see the human at the centre of the interaction, with all her virtues and, perhaps more importantly, limitations. We humans are not machines, and thus require the tools and artefacts we use to be very well adapted to our needs and prerequisites. With this perspective on interaction, one often finds many systems to be lacking in several areas such as reliability, trustworthiness or complexity.

One such aspect of contemporary HCI is temporality. Temporality is the intrinsic relationship between a something and time. In HCI, this would mean the relationship between a user interacting with technology, and it responding in a timely manner. We humans do, after all, live in a world where time is of the essence. Nobody lives forever, and our minds are shaped by a world where time flows at a constant pace, no matter what we do. It should therefore not be a surprise to anyone that temporality is an important part of every-day life, including of course, computer interaction.

Temporality in interaction has been on the minds of researchers and developers alike for over 40 years, and yet it remains a largely uncharted aspect of interaction (see Hoxmeier & DiCesare, 2000; Huang & Stolterman, 2011; Dabrowski & Munson, 2011). The field lacks definitive answers to questions regarding how sensitive humans really are to delays, and to what the effects of temporality are on human cognition, to name just two of many possible questions.

What is accepted, however, is the general importance of temporality in interactive systems. Many studies have shown that temporality is a crucial part of computer interaction (e.g. Barber & Lukas, 1983; Shneiderman, 1984; MacKenzie & Ware, 1993; Pantel & Wolf, 2002) and that prolonged delays cause user frustration and a decrease in performance in certain tasks. System responsiveness, or how well user actions are coupled with system response, has been shown to be the most important factor when determining user satisfaction (Johnson, 2010, p. 151; Rushinek & Rushinek, 1986).
Despite system responsiveness being widely accepted as a major contributing factor for user satisfaction and perceived speed of use – at least within the scientific community – many of today’s systems suffer in this department (Johnson, 2007, in Johnson 2010, p. 153; Roast, 1998; Huang & Stolterman, 2011). In particular, the current trend to move away from traditional desktop systems, and into the ”cloud” has brought with it temporal uncertainty in interaction, beyond the control of either the developer, or the user. Temporal uncertainty is the phenomenon where the user cannot predict a timely response, due to, in this case, varying response times from servers, or other internet peers. This is an especially alarming development, considering the importance of temporality and timely feedback, and could grow to become a prevalent tribulation of computer interaction in the near future. It already is to some degree (see Galletta et al., 2006; Dabrowski & Munson, 2011).

With an industry striving to achieve the best possible user satisfaction, it seems strange that the temporal aspects of system design are not given more focus. For this to change, however, there is a need to further our understanding of these underlying issues, as well as to shed more light on the importance of these aspects of interaction. The potential benefits of expanding our knowledge in this area are noteworthy. A better understanding of human-computer interaction – especially in terms of temporality – would allow software developers to create better systems, more fit for their human users. Better systems, and subsequently less annoyed users, would mean a better world for everyone – reason enough for fields such as HCI to exist, and for this study to be of interest and importance to the academic field and industry alike.

Thus, the aim of the current study is to examine the relationship between the human operator on the one hand, and temporality on the other. An experiment tests how sub-second delays affect users in simple, but cognitively demanding, computer interaction. It is hypothesized that even delays as short as a few hundred milliseconds will have a significant negative effect on users. A general theoretical background introduces the field and creates a context in which the conducted experiment can be situated.

1.1 Purpose

The main purpose of this thesis is to consolidate the existing research that exists within the fields of system delays and responsiveness, to create a comprehensive and coherent summary of the subject. The purpose being, to further the development and expansion of interactive systems and human computer interaction, and to facilitate a thorough understanding of the existing issues in the field.
The results of an experiment designed to test the effects of sub-second delays on computer interaction are also reported. Sub-second delays were chosen, as the contemporary research appears to be insufficient in that particular area and is therefore of interest of investigation.

The intended reader is either an academic (within the fields of human-computer interaction) or a developer in the industry (such as a designer or a software engineer) Research questions were therefore chosen for their academic merits as well as for their value to the industry.

1.2 Research questions

The main research questions addressed in the current study are as follows:

- What are the effects of system delays on users operating an interactive system?
- Does this effect differ depending on the type of interactive system?
- Are there quantifiable temporal 'deadlines' by which different types of systems must adhere?
- Are there any positive effects of delays?
- Can delays be anything more than simple 'wait time' for users?
- Do constant, sub-second system delays affect user performance in a simple computer program?

1.3 Research approach

To answer the some of the research questions, and to fulfil the purpose of consolidating the current understanding of the subject at hand, there was need to review the last 40 years of literature in the field of temporality in interaction. It was of importance to first introduce and define central concepts to interaction, temporality as well as categories of systems, before situating previous research. With such definitions complete, it was possible to produce a consolidation of current research in a review of literature, encompassing the field of temporality in interaction, and further categorise the chosen studies by, and into, the a priori defined categories. This review has been made broader and more extensive, than would be required for the conducted experiment, as a deeper understanding of the underlying issues and their context is needed, and indeed valued. From this review, an area of academic deficit was identified, where an experiment could further add to the current understanding of a specific issue.
1.4 Delimitations

The consolidation of previous research is bound by the *a priori* definitions, and further categorisations of the chosen articles. This approach assumes a certain perspective on interaction in favour of some aspects, but to the detriment of others. The current study concerns itself with the interaction between man and machine (HCI) and does not take into account the practical limitations of system design, for example.

This study uses an experiment that examines sub-second delays for a specific type of computer interaction. There is ample reason to believe that the type of interaction has a large impact on the effects of temporality (e.g. Kohlisch & Kuhmann, 1997), which would mean that the results from the experiment are probably only applicable to that very same type of interaction. The experiment also only examines constant system delays, when variable delays are perhaps more common, and can have different effects on users entirely. However, there is at least some form of translatability, which allows for some general assertions to be made from the results, especially in conjunction with existing theories.
2 Background

This section will introduce the favoured terminology, and establish the chosen paradigm for human-computer interaction. Furthermore, a review of literature is presented which contains a selection of the major previous works within the field. This is concluded with a summary of the assertions made from the literature. The last section introduces the rationale behind the conducted experiment and presents the hypothesis in further detail.

2.1 The foundation of this study

This section will define the chosen perspective on the human operator, a distinction between two different kinds of interactive systems, and lastly, the temporal aspects of interaction. These definitions constitute the foundation of the current study and will shape the further analysis and discussion.

2.1.1 The human operator

In the scientific fields of HCI and human factors, the human operator is at the center of the interaction. These fields view the human as the one in control, and regard the interactivity between said human, and their artefacts, as paramount. The current study is based on the assumption that this stance on computer interaction is correct, and uses these paradigms in both the review of literature as well as in further analysis and discussion. The term user will henceforth be used to describe any human operator, so to reduce redundancy.

To define the user, and his/her many possible stipulations, cognitive or otherwise, the Contextual Control Model (COCOM) is appropriate. COCOM was developed by Hollnagel and Woods (2005), originally to model joint cognitive systems (JCS), or other complex systems, but can also be used to describe and illustrate the user and how he/she interacts with a computer, or other digital artefact.

COCOM, as illustrated in Figure 1, shows how the user chooses an action and receives feedback from the system, which in turn is used to choose the next action. This repeats itself in a cyclical manner, and represents how the user can maintain control of a system. It demonstrates that adequate feedback, as well as knowledge of what actions are available, is necessary for the user to maintain control. Within cognitive systems engineering (the field in which COCOM originally belongs) maintaining higher control diversity than the variety of the system is key, as per the Law of Requisite Variety (see Hollnagel & Woods, 2005, p. 40). In the case of this study, and HCI in
general, COCOM can be used to illustrate the need for *timely* feedback, in the same way that feedback in general is necessary.

Figure 1: COCOM for interactive systems

When a user initiates an action, he/she wants confirmation of its initiation, as well as of its completion. The user may not be able to choose the next action if the previous has not produced any perceivable results (e.g. the system is too slow). Figure 1 also shows the difference between feedback on initiation of action, and feedback on the true results: ”*Confirms*” and ”*Results*”, respectively. It is not the same thing for a system to confirm an action, as it is to actually produce the results of said action. Confirmation of action needs to happen almost instantaneously, so the perceived causality between action and system re-action is maintained (Johnson, 2010, p. 160).

### 2.1.2 Two types of interactive systems

A distinction can be made between two types of interactive systems, and is based on a fundamental difference in how users control an interactive system. A user can maintain control by using either discrete, or continuous actions – discrete in the sense that the available actions are distinctly separated from one another, and likewise, continuous in the sense that the actions lack discrete steps.

Thus, in the current study, these two systems will be called discretionary and continuous, respectively, so as to describe their defining trait. Discretionary
and continuous systems have a certain similarity to open and closed loop control. Open loop control means that the user disregards system feedback, and chooses actions entirely based on the current state of the system, as well as his/her own mental model of said system. This is, to a certain degree, similar to a discretionary system, as the user goes from one discrete state to another, choosing new actions. Obviously, no user is ever truly operating entirely by open loop, as the resulting interaction would be very inefficient. Such control would also contradict the chosen model for the user (COCOM) as it clearly goes against the cyclical nature of the user paradigm. Thus, feedback is always going to be relevant for the user, if only to know where the user has been, and where he/she is going.

A discretionary system, on the one hand, is an open loop system in the sense that its states and actions are discrete. These states and actions are not, however, entirely detached from previous states, as the user still requires feedback from prior actions to choose the next ones. Browsing the web is a clear example of a discretionary system, as the user goes from one page to another, but not without regard to his/her previous steps.

A continuous system, on the other hand, is more true to a closed loop system. The user requires good feedback from previous actions, as maintaining control relies heavily on the results of prior actions. Merely a mental model of the system is insufficient, as these types of systems often operate in domains characterised by their stochastic nature. Moreover, such a system often requires continuous control, in the sense that a user cannot "let go", and has to continually feed the system actions or else it stops functioning properly. A good example of a continuous system is the task of driving a car. Maintaining control of a car requires fine motor control, among many things, which can only be achieved by knowing the results of previous actions – i.e. good feedback. Furthermore, turning the steering wheel has no discrete states, and illustrates how continuous systems can operate.

Moreover, temporal constraints are also likely to be more prevalent with continuous systems than with discretionary systems. The more a system relies on closed loop control, the more important timely feedback becomes.

This study is not the first to have made the distinction between these two types of interactive systems, as others have previously had similar ideas. Dabrowski and Munson (2011) made a distinction between what were essentially discretionary and continuous systems, only they chose to call them conversational systems and control tasks, respectively. In their article "40 Years of searching for the best computer system response time" they present a review of literature, not unlike the one in further down in this chapter, and propose a framework for how systems should be regarded when designing for temporality, following their own proposed distinction. Their article can be used as an alternative, albeit similar, perspective on temporality in
interaction, as they have structured their work based on the type of effect the delay is thought to have, rather than the type of system within which the delay occurs.

2.1.3 Defining temporality

Finally, the most important definition is the term temporality, and how it fits in with our prospective user, and the two types of interactive systems. Temporality is elusive, both as a concept for us human beings, and as a term within HCI. Temporality can take many forms, and whilst most are important to the advancement of HCI, and necessary for the betterment of today’s interactive systems, discussing them all would be outside of the scope of this study. The current study concerns itself with the temporal aspects of immediate interaction, i.e. the interaction that happens between the human and computer, in the moment. With regard to the previously defined model of the user (COCOM; see 2.1.1), one such interactive cycle can last from basically no time at all, to a few minutes. Of course, the whole purpose of having a cyclical model is to illustrate the recursive manner in which interaction occurs, meaning that the interaction is often not finished after merely one cycle.

![Figure 2: The temporal aspects of COCOM](image)

Although there is more to temporality in interaction beyond just a few minutes, this study does not include those parts. As interaction goes beyond a few minutes it changes in nature, and becomes much more unpredictable (at least seen to user behaviour) (cf. Huang & Stolterman, 2011).
Thus, the defined interaction governs a user’s interaction with a system where each action must have a relatively rapid system response (see Figure 2; confirmation of action initiation). This intrinsic relationship creates and defines the term responsiveness. Of more importance however, is the definition of the time-period between a user action, and the system re-action (Figure 2; production of results). Different fields, scientific or otherwise, have different terminology for this temporal aspect. In visual representation, for example, we find latency as well as lag being the most frequently used terms to describe the phenomenon. In more traditional computer science we often find terms such as system response time, or system delays being the most widely used. Although a difference exists between the terms latency, lag, system response time, etc., they should be considered synonymous, within the confines of the definitions provided.

The term system delay will henceforth be used to describe this phenomenon, so that consistency is maintained throughout the study. In the review of literature below, however, the original terms used by the authors are maintained so as not to misrepresent their findings.

2.2 Review of literature

In the following review of literature, a selection of publications in the field of temporality in interaction are presented and discussed so as to summarise the current understanding for some of the most prominent issues in the field. The review, and the research presented within, is categorised based on a general theoretical framework and further the type of system and task being studied, and is written in a semi-chronological order. The review is concluded with a summary, where assertions made from the literature are presented.

2.2.1 General theoretical framework

Temporality has been a noteworthy scientific aspect of computer interaction since the inception and spread of computers, now over forty years ago. The scientific field has come a long way since then, mainly due to advancements in computer science and in the field of the temporal aspects in interaction. However, the temporal limitations of user-controlled systems have been known for longer than that. During the 1960’s, tele-operation was considered as an alternative to sending humans into space. However, placing humans far from the thing they intend to control in space, also meant separating the two in time, i.e. introducing long transmission delays between operator and machine.
A study by Sheridan and Ferrell (1963) examined the effects of transmission delay on remote manipulation tasks. The virtues of remote manipulation are many and go beyond mere space-flight; to be able to explore the deep ocean floor and to carry out difficult tasks in hazardous environments – are among the many things remote manipulation excels at. These environments require the cognitive abilities of humans, but does not support the presence of their physical bodies, and so require some form of remote manipulation. Remote manipulation is furthermore contrasted with continuous control situations, as remote manipulation lacks the immediate (as a function of time) tracking of continuous control situations. Remote manipulation is more about “modifying certain initial positions of objects in the environment to achieve given final positions” (p. 25). The authors conclude that operators consistently use an open-loop approach along with a wait-and-see strategy in remote manipulation tasks with transmission delay. Operators maintained a stable control throughout the conducted experiment, and increased time delays were shown to only gradually increase completion time for tasks.

However, it was not until the 80’s that system delays started to see more scientific exposure. Many early studies wanted to know if prolonged system delays affected the users in more ways than just their task performance. A study by Rushinek and Rushinek (1986) found that good system response times were the most important factor for determining user satisfaction. They further noted that users did not like to wait for the system to respond. Another study by Shneiderman (1984) provides evidence that system response time has an effect on error rate and user satisfaction, beyond just user performance. Shneiderman (1984) did, however, find that error rates increased when response times were too short, and argues that a ‘reasonable system delay’ (in this case 12 seconds) is preferred, as it improved error-rates. Rushinek and Rushinek (1986) and Shneiderman (1984), both showed the importance of temporality in interaction, when the field was still relatively young. They did not, however, discuss how these findings would differ in other situations other than the ones studied, and were generally rather moderate in any practical implications presented.

A study by Barber and Lucas (1983) employed an experiment where system delays were introduced onto unsuspecting computer users operating a real transaction system. Neither the users, nor their supervisors, were aware of the experimental condition – the users were basically doing their job as usual. System delays ranging from the system’s baseline average of six seconds, and up to 14 seconds, were introduced, and their implications measured. The study was able to provide further evidence that system delays decrease both operator performance and job satisfaction. Error-rate increased non-linearly, but were at their lowest at a system delay of 12 seconds, creating a U-shape across the measured delay times. This was one of the first studies on temporality conducted in a naturalistic setting, and
thus gave the researchers a good idea how delays affected people in real world situations. Like Shneiderman (1984), Barber and Lucas (1983) support the notion that error rates are lower when there is a moderate system delay.

Not all studies were able to show the negative effects of prolonged system delays. There were a few failed attempts of finding measurable effects (see Dannenbring, 1983; Butler, 1983; Dannenbring, 1984) and are often found in, and likewise limited to, the early 1980’s. One has to wonder why, during this time-period, evidence for long system delays affecting user performance was so elusive. There seem to be several reasons to why this is the case.

Teal and Rudnicky (1992) mention these issues in their article about how researchers have failed to adequately address the issue of system delays. The authors propose a few different reasons to why previous studies might have failed to produce measurable effects of system delays. Choice of task, amount of user control within each task as well as over the delay condition, and also the match between delay condition are a few among many of the possible explanations presented as to why different studies have had difficulties of finding conclusive results in the past (p. 296). The authors further argue that these previous inconsistencies have resulted in a lack of clear and precise guidelines to designers and engineers to use of in making crucial design decisions in their work. They further request the development of a predictive engineering model governing user performance and system delays, and propose their own such model, by examining how different system delays affect user strategy selection. The authors define three types of user strategies and hypothesize that at certain times the cost for each interaction strategy exceeds a threshold and that users subsequently changes to the next type of strategy. These strategies correspond to a certain amount of system delay, where users go from exerting continuous control at non existing or short delays, and transition to a more open loop, wait-and-see strategy when delays are prolonged. This transition away from continuous control as the delay increase, can be seen in Sheridan and Ferrell (1963), described above. The long distance between operator and artefact forces the operators to use a more open-loop, wait-and-see strategy.

The definition of a strategy selection model dependent on system delay was by no means undisputed. A short article by Johnson and Gray (1996) describes a group of researchers trying to replicate the results from Teal and Rudnicky (1992). They did not, however, succeed in doing so, and argue that the failure to reproduce the results are most likely due to the problems that arise when different research teams are forced to use unclear or imprecise descriptions.

Likewise, O’Donnell and Draper (1996) contested the findings of Teal and Rudnicky (1992) in another attempt to replicate the results. The authors
presents evidence that user strategy selection depends on delay length. The authors are not, however, entirely convinced by the idea that this is solely determined by system delays, and argue that there are many more aspects to the choosing of strategies than merely delay length. They further observed that users did not commit entirely to one of the strategies defined in Teal and Rudnicky (1992), but tended to switch between several strategies. The authors argue that the choice of strategy is much more clear in introspection than it is by looking at only response times and error-rates. This is a problem they think should be addressed before such an experiment could be seen as definitive evidence that a certain system delay causes the user to re-evaluate their strategy. Furthermore, they make the point that users do not work like machines, but take small breaks, adjust their posture and double check their work before continuing on – things any person would do in a naturalistic setting. Thus, a certain system delay causing specific user behaviour is highly improbable (p. 41). The choosing of a strategy is far too subjective, and likewise does not constitute as a big commitment from the user, as they can change strategies continuously. The authors conclude the article by the observation that research in this field is a long way from producing recommendations for designers and software engineers, again echoing the request of Teal and Rudnicky (1992) for a more rigid theoretical framework.

Kohlisch and Kuhmann (1997) examined if there is such a thing as an optimal inter-task delay, i.e. the system delay in-between tasks. They did so by measuring skin conductance, heart frequency, error rate and irrelevant keystrokes, in situations with varying system delays. The authors found that all these measures pointed to a 5 second delay between tasks being the optimal for user performance and user well being. They do argue, however, that this optimal time is task specific, and needs to be studied for each task independently. In conclusion, the authors argue that medium inter-task delays are optimal, and what counts as medium is highly dependent on the task. It is generally hypothesized that more complex tasks sport longer medium delays. Nonetheless, the article provides a counterpoint to the idea that lower system delays are always better, with regard to user performance. The notion that there exists an optimal delay, which is not the lowest possible, was at the time of publication, a novel conception. The existence of an optimal inter-task delay can explain the U-shaped error-rates of Shneiderman (1984) as well as Barber and Lucas (1983). For example, Barber and Lucas (1983) showed that the best delay was 12 seconds (with regard to user performance), which indicates that the type of task being performed affects what constitutes as an optimal system delay.

It seems likely that system delays have an effect on error rates, beyond user performance and well-being. But why do medium system delays seem to result in better error-rates? A possible explanation is that the type of task
performed (often manual text/numerical input) affects the users strategy selection (see Teal & Rudnicky, 1992) and that users choose to operate the system in a different way when the delay is of a medium length, than when the system is as quick as possible. As to why Kohlisch and Kuhmann (1997) reported improvements on many other aspects, other than just error-rate, Szameitat et al. (2009) provides a commentary on many previous works (including Kohlisch & Kuhmann, 1997) on why longer system delays could result in positive user performance. They argue that the effects of a delay is highly dependent on where (or rather when) in the task it is presented. Delays in between tasks are not equally detrimental to users as delays that actually interrupt the interaction. Delays between tasks or subtasks could give users more time to prepare for the next sequence of actions, and is thought to be the reason why some earlier studies have showed increases in performance after increased system delays (see Kohlisch & Kuhmann, 1997; Shneiderman, 1984; Barber & Lucas, 1983).

Szameitat et al. (2009) present the results of an experiment, where delays ranging from 500 msec to 2.8 seconds were applied to a simple, discretionary, computer game and where the delays were presented in random order, and only some of the times an action was initiated (this was done in order to minimize user anticipation of delays). These delays were shown to have a negative effect on users, as the error rates more than doubled following a system delay. The authors provide evidence that system delays as short as a few hundred milliseconds can still have considerable detrimental effects on performance and emotional well-being, if the delays constitute as interruptions to users. Given these results, it seems more than likely that not only delay length affects users, but perhaps more importantly; where and when in a task it is presented (see Kohlisch & Kuhmann, 1997).

So far the field has seen studies conducted on both the length of system delays, as well as where and when during a task they are presented. During the last few years, however, another aspect of temporality in interaction has blossomed in the field of HCI. Thomaschke et al. (2011) examined how temporal expectancy affects response to stimuli – essentially studying predictability. The authors use the foreperiod paradigm, where the imperative stimulus is preceded by a warning stimulus. The foreperiod paradigm can be described as "[w]hen participants expect that the imperative stimulus will occur after a specific [foreperiod], they can prepare for the imperative stimulus, and, consequently, respond to the stimulus more quickly" (p. 2309).

The foreperiod paradigm concerns the period in which a user has to wait for a stimulus, and does not include such things as continuous feedback (e.g. a progress bar), but a period where stimulus and feedback is absent. Predictability comes from the consistency of foreperiod length, rather than continuous feedback to the user. A system is therefore predictable when the foreperiod is of consistent length, and a prospective user can get a feel
for, and get used to, the delay length after continuous use. The authors hypothesize that specific temporal expectancy increases the excitability of the spinal cord, which in turn could explain why user performance (in this case reaction times) are improved after a certain constant foreperiod stimulus. This basically means that variability of system delays have a detrimental effect on user performance. High temporal expectancy helps users adapt to a certain delay, potentially neglecting its original detrimental effects. Low temporal expectancy – i.e. variability of system delays – would therefore hinder user adaptation. This is something Thomaschke and Dreisbach (2013) give further evidence of in their study. Rolke and Hofmann (2007) also show support for the idea that temporal uncertainty has negative effects on users as they showed that user performance was enhanced when predictability was high. Rolke and Hofmann do maintain that predictability in interaction is an important factor when determining user performance, and should not be neglected.

A study by Thomaschke and Haering (2014) present the concept of predictivity of system delays. Predictivity is not the same as predictability. "Predictivity and predictability are inverse concepts: predictive delays predict, and predictable delays are predicted" (p. 360). Predictivity of delays is related to which system response comes after a certain delay [length]. A delay is predictive if it indicates what the system response is going to be (i.e. 1 second system delay for an error message, and longer delay for successful results). A user can therefore know if their request has been successfully inputted if the system always shows an error message within the first few seconds, and if their own request has buffered longer than that. Thomaschke and Haering (2014) conclude that deterministic predictivity (where the probability of a certain result after a given delay is 1) outperform all other conditions; constant and variable system delays. Wagener and Hoffmann (2010) further supports this statement, and argue that the point in time which something occurs, will prime the processes which are typically to perform at that point in time.

The aspects of predictability and predictivity give an alternative perspective on temporality in interaction, as system delays can be much more than just a set length of time. Delays can even be used as feedback to the user, and so assist and alleviate in their work (see Thomaschke & Haering, 2014).

Branaghan and Sanchez (2009) present an article on different types of feedback when there is delay – ranging from progress bars to more static loading screens – and what type of effect these have on both the user’s perception of the delays as well as the user’s attitude towards the type of feedback used. The results showed that users generally like more information, that is, more predictability from their feedback, but at the same time perceive the waiting times as longer the more descriptive and deterministic this feedback is.
The authors attribute this effect to the human mind’s innate ability to sense the passing of time. The more distinct temporal cues the user receives, the longer that same time period will feel to a user. The authors argue that there is a crucial balance towards perceived speed of use and satisfactory feedback, obtained through creating as few distinct events as possible in the feedback, while at the same time providing coherent information about the remaining time to the users. These findings stand in contrast to previous studies on predictability and indicate how high predictability is not necessarily always a good thing.

Until only recently, studies have merely presented evidence for individual effects on specific types of systems or tasks, but not taken any further steps to generalise these results. Any general theories, or encompassing theoretical frameworks, have been conspicuously absent (cf. Roast, 1998), albeit frequently requested (e.g. Teal & Rudnicky, 1992; Kohlisch & Kuhmann, 1997). Dabrowski and Munson (2011) try to answer to the general lack of theories, by proposing their own such theoretical framework. They begin by summarising the last 40 years of studies on what they call “the search for the best computer system response time”. They further present a model that divides systems into two categories; conversational tasks and control tasks. This distinction bears a clear resemblance to the distinction between discretionary systems and continuous systems presented in the current study. Conversational tasks (derived from the conversational nature of these tasks), they argue, require some delay to maintain a natural and productive pacing of the system. Furthermore, Dabrowski and Munson (2011) (like Kohlisch & Kuhmann, 1997; Szameitat et al., 2009), argue that the importance of location and duration of delays depend largely on task complexity (i.e. type of task) and user expectations. Control tasks, however, perform best at immediate response times, according to the model presented.

This theoretical framework was, at the time of publication, one of the first studies to create a comprehensive summary of the field of temporality in interaction, and create a general theory for temporality along with system response time guidelines. This gives designers and academics alike some basic guidelines in the further development of the field, something the authors themselves request. While their framework might be comprehensive and provide a necessary perspective on the field, there are still many unanswered questions, especially when it comes to task-specific sensitivity to delay within conversational systems (discretionary systems).

There also exist frameworks for temporality outside of the academic world, for example the one presented in Jeff Johnson’s book ”Designing with the Mind in Mind” (2010). Johnson (2010) makes a similar distinction to Dabrowski and Munson (2011), between the perceptual-motor and automatic processes [of control tasks] and conversational tasks. Johnson (2010)
proposes a temporal framework that builds upon the simple assumption that systems should not be perceived as lagging behind (i.e. being unresponsive) by the user. In this framework, he maintains that control tasks need to respond immediately (defined as shorter than 0.1 seconds) to user interaction, as otherwise the causality of actions disappears. Conversational tasks, on the other hand, Johnson (2010) argues should not have a delay of more than 1 second. These two deadlines, along with all the others presented in his temporal guidelines, are derived from time-scales applicable to the human brain and its functions. For example, conversational tasks should not have a delay of longer than 1 second, because gaps in normal human conversation often do not exceed this length (p. 155). This constitutes the one major critique that can be had towards Johnson’s (2010) framework, in that the proposed guidelines are not based on empirical evidence of actual results and performance effects of delays, but rather extrapolations and assumptions from neuroscience and psychology, on how fast our brains process certain types of information. However useful his guidelines may be, the fact remains that they fail to address many of the contemporary issues in the field, as many specifics of temporality are omitted. Despite the omission, Johnson (2010) recognises several key aspects of what creates a responsive system, and by doing so, helps shed light on the importance of system responsiveness.

2.2.2 Discretionary systems

When it comes to discretionary systems, many, if not all, of the studies described above apply, because they are both general as well as fit into the category of discretionary systems. There have, however, been studies conducted on areas that examined the use of specific discretionary interactive systems, and were therefore better categorised in their own subsection. Use of the web, and tolerance of delays on the Internet, will be the main focus of this category, not because it is the only type of discretionary system, but because it is one of today’s most commonly used discretionary systems, and an area that appears to have inspired interest amongst many researchers.

A study by Hoxmeier and DiCesare (2000) examined how tolerant web users were to delays. They found that user satisfaction decreased with increasing response times. They also found that the threshold for when a user abandons a website seems to be at around 12 seconds of delays. It is suggested that too short response times decrease user thinking time, which can cause users to interact with the system at a faster pace, which in turn could lead to an increased error rate. One of the limitations of this study is that the participant were not under any time pressure when performing the different tasks, which could have affected the results – user could be either more or less patient with a system than they were in the experiment.
A study by Selvidge et al. (2002) found evidence that faster loading times were generally better for user satisfaction and performance, with an exception in which users felt a sense of being lost when the response times were too short. The authors propose (in accordance with previously mentioned studies, such as Kohlisch & Kuhmann, 1997; Szameitat et al., 2009) that the type of task should dictate how long the response time should be, and maintain that some tasks could benefit from longer, rather than shorter, delays. Galletta et al. (2004) provides further evidence that user performance starts to degrade gradually after only a few seconds of delay during web-use. User attitudes were also shown to be affected by longer delays. However decisive these previous findings may seem, one should not forget these studies were conducted often more than a decade ago. The internet, use of the web, and user online habits have changed dramatically over the years, and should stand in contrast to these previous studies.

A slightly more modern study by Galletta et al. (2006) examined how website delay, in combination with site familiarity and site breadth, affects user performance and satisfaction. The study found that these other two aspects of surfing the web could help reduce the negative effects of long delays. The authors argue that both familiarity and site breadth should not be studied independently, but seen as complimentary to the system as a whole. This could be seen as an indication that the negative effects of system delays can be reduced, or entirely mitigated, by other aspects of a system. Whether system designers should actually rely on this capacity remains in question, however.

Furthermore, Galletta et al. (2006) highlight the problem of variability and temporal uncertainty in web interaction. Stating that "The ultimate speed of loading a page is determined by the weakest link in the transmission path from the server [...] to the user’s desktop" (p. 21) and further notes that "Even without congestion, users can still find themselves in a global waiting line at a popular site, or the victim of server configuration errors" (p. 21). The authors illustrate that temporal uncertainty can become a prominent issue in contemporary computer interaction, and to some degree already is (Dabrowski & Munson, 2011), and argue that this is something developers should keep in mind when designing system on the web, given the detrimental effects of variable delays (see Thomaschke & Haering, 2014; Wagener & Hoffmann, 2010).

Sellier and Chattopadhyay (2009) studied user motivation in online goal pursuit where different lengths of system delays were introduced with users performing different tasks – in the form of task interruptions. The authors found evidence that motivation increased when download times were moderate (a few seconds), contrary to systems harbouring shorter download times (0-1 seconds) which did not result in the same positive effect. They show
that, in some cases, longer rather than shorter interruptions helped increase motivation to continue pursuing a goal. Motivation, however, is not akin to user performance; it just shows that users are more likely to continue with their task if the interruption was conceivably longer than a blink of an eye. This does also support some of the previously mentioned studies showing that moderate rather than short system delays have positive effects, but goes against the current theory (see Szameitat et al., 2009; Kohlisch & Kuhlmann, 1997) on why moderate delays are better. Most likely, what the study measured was not true interruptions, but rather inter-task delays, where the users had time to recuperate and think before proceeding with their next step, pursuing their online goal.

Contrary to the results of Sellier & Chattopadhyay (2009) and Hoxmeier and DiCesare (2000), Nah (2004) showed that users were only willing to wait approximately 2 seconds for a website to load (in an information retrieval task), before choosing a different site, or opting out entirely. The existence of feedback was shown to prolong this time somewhat. The authors, however, fail to discuss the placement of delays (were they interruptions, or inter-task delays) in their experiment, but do mention that the approximation of 2 seconds maximum delay may vary for different types of tasks. Generally, many studies seem to examine and discuss the same underlying issue, but may in fact be looking at two entirely different type of delays. Until these aspects get a more prominent role in this type of research, we will have to content ourselves with imprecise definitions and tentative time limits.

2.2.3 Continuous systems

Continuous systems are, contrary to discretionary systems, much more control dependent, and thus much more sensitive to temporal disruptions. These types of systems often work under very specific circumstances and require their own, tailored, set of guidelines, why the general theoretical framework may not apply.

Teleoperation is one of the many examples of a continuous system, and has been on the minds of researchers since the inception of the field of temporality in interaction (see Sheridan & Ferrell, 1963), and can be defined as the remote control of machines. A study by Thompson et al. (1999) studied the effects of time delay and asynchrony in telesurgery; the teleoperation of surgical equipment. The article describes an experimental condition where a surgeon is controlling a remote surgery machine, operating on a patient together with an assistant located on site (it was not if the patient was an actual person or not). Different experimental situations were created to examine the effects of different types of task division between off-location surgeon and on-location assistant, as well as asynchronous and synchronous
feedback delay in these conditions. The study found no clear thresholds for delays, but did notice a performance increase with asynchronous video-tactile feedback. The authors hypothesize that the tactile feedback was sufficient when performing the required tasks, and that slowing the process down to fit the video feedback (as was necessary with limited bandwidth) led to worse performance overall. As the study failed to find evidence of the effects of delays, and because the tasks were of a simple nature, the results are more useful as a showcase for some of the issues that can arise in telesurgery, rather than for providing any clear guidelines for the temporal aspects in such an application.

Another application of teleoperation exists in space-flight. Lester and Thronson (2011) examined how latency affects telerobotics in human space exploration. The term “cognitive scale of the universe” is introduced, and proposes the distance a human operator can be from an operation he/she wants to control, and still maintain unimpaired cognitive abilities. They offer a discussion about human space-flight, and how future missions to the unknown can be performed. Sending humans in the flesh to space is an extremely complicated, dangerous, and not to mention, expensive endeavour. Sending only their cognitive abilities is (often) a more viable option, when technology offers teleoperation with equivalent control and perceptive abilities for the operators. The authors argue that that "when the latency is sufficiently low and telepresence sufficiently sophisticated, the experience of an astronaut on-site, encased within a constraining and sense-dulling spacesuit, is likely to be far less ‘real’ than that of humans immersed in high-quality telepresence" (p. 90). They further note the limitations of latency in space exploration, as it is the speed of light that keeps teleoperation in space unrealised, giving an example that a two-way communication to the Moon has a minimum latency of about 2.6 seconds, assuming all other lag-inducers are removed.

Lester and Thronson (2011) conclude that the amount of tolerable latency in teleoperation depends strongly on exactly what one is trying to accomplish. They present an example of how operators are able to control the Voyager spacecraft, now more than a day’s latency away. It should be noted that the control of this space-craft has long since changed from direct control, to a wait-and see approach, like that of the rovers on some of our closest celestial object, the Moon and Mars. Lester and Thronson (2011) further propose that latency should not be greater that 300-400 msec, after which important cognitive and motor-perceptual abilities are reduced. The practice of teleoperation and success of missions will still be limited by such earthly things as the human factor, and specifically the operator’s ability to maintain control in high latency situations, which is a humbling realization when dealing with such alien issues as space exploration.
A study by Gawron et al. (2010) examined the effects of delays in cockpit simulation tasks as well as in real aircraft – essentially examining how robust the closed-loop pilot-vehicle system is to temporal disruptions. The authors observed that performance degradation was greater in the simulated scenario than in the real in-flight situation. From the full motion flight simulation, 150 msec delay from cockpit input to visual response was shown to be the maximum allowable delay before flying qualities degrade from Level 1 to Level 2. Levels are a standardized measure within aeronautics in measuring flying qualities. The three levels are (1) satisfactory, (2) acceptable and (3) controllable. The authors further note that a reasonable design criterion for flight simulations would be that delay no longer than 50 msec be allowed. The authors predict that any longer delay would degrade pilot performance, or otherwise increase pilot compensation, more than would be the case in actual flight. As simulations are trying to replicate the conditions of the “real thing”, it is important that issues such as delays do not effect the users in any significant ways, thus lowering the authenticity of the simulation, and further dislodging the pilot from the illusion of actually flying an aircraft.

A different perspective on the issue of delays in continuous systems is found in Chen et al. (2007), where the authors present a framework for the design of teleoperation interfaces (applicable to issues presented in Gawron et al., 2010; Thompson et al., 1999 etc.) and further propose a solution to the problem of asynchronous/delayed teleoperation. They suggest that designers create an interface that predicts the results of the executed actions in real time, before the actual return signal has arrived. This would mean that the screen/simulation acts as if there was no delay in the interaction, predicting what the consequences the actions would have on the controlled object/environment, effectively surpassing the wait-time for the “production of results” (see section 2.1.3, Figure 2). When the actual feedback does arrive, the system has to correct its prediction, comparing what it thought would happen after a given action, and what actually happened. Our brains are actually already doing this very thing for our vision, as it takes about 100 msec for our visual cortex to process visual stimuli (Stafford & Webb, 2005, in Johnson, 2010). Chen et al. (2007) further present evidence that users controlling a driving task through teleoperation have a limit of about 170 ms delay before performance is decreased. They argue that this threshold is highly task specific, and that other tasks require their own specific limits.

A different type of continuous systems is that found in (online/multi-player) games, and has been the subject of much research. Pantel and Wolf (2002) examined how players of a driving game were affected by increased system delay. The study concluded that delays of up to 50 msec were hardly noticeable and did not affect the performance of the players. At 100 msec delay, there was a noticeable performance decrease, and at 500 msec the game
became completely unplayable. These values should hold true for any continuous control task that requires fine control over a fast moving object, in a two-dimensional context. MacKenzie and Ware (1993) present similar findings in their study of two-dimensional hand-eye coordination tracking tasks, in that lag was easily noticeable at 75 msec, and caused severe problems above and beyond 225 msec.

Beigbeder et al. (2004) studied how packet loss and latencies affect user performance when playing the game Unreal tournament 2003. They showed that packet loss, as severe as 5% had no measurable effect on user performance, whilst latencies above and beyond 100 msec had a significant effect on player performance. They argue that multi-player games should opt to use a faster mechanism in favour of lower latencies, rather than a more secure data packet transfer mechanism. Furthermore, the study provides evidence that latencies continue to have an inverse relationship with user performance, and that latency of 200 msec and longer were considered annoying by the players.

Yet another type of continuous system is Virtual Reality (VR). A study by Allison et al. (2001) examined how tracking latency in VR environments affects users. The general assertion made by the authors is that latency is bad for user performance as well as comfort. Because VR – usually induced via a headset – covered large parts, if not the whole, of a person’s field of view, display latency is both more noticeable, as well as more detrimental, to user performance and well being. Ellis et al. (2004) provides evidence that latencies of about 10-20 msec can be noticed in a VR environment. The authors maintain that latencies in VR, or virtual environments covering large parts of a user’s field of view, need to be kept below 20 msec for optimal results.

There are, of course, many more applications of continuous systems beyond the few mentioned and discussed here. It is beyond the scope of this study to touch upon and discuss them all, and those presented should be sufficient to convey the general temporal limitations of continuous systems. Generally, however, temporal limitations of continuous systems are very much limited by the type of system to which it relates. Continuous systems are, because of the type of control exerted by the user in such systems, much more sensitive to delays than are any discretionary system, an exception being any system that is controlled by a wait-and-see strategy. Such a system would arguably not be regarded as a continuous system (as per the definitions, see section 2.1.2), but instead fit in the realm of discretionary systems. Strategy selection in continuous systems should also function the way described and discussed in Teal and Rudnicky (1992), in that when the cost for maintaining a certain strategy exceeds a threshold, the user changes to a looser manner of control.
2.2.4 Summary of reviewed literature

There are several points worth taking away from the above review of literature. There are no definitive answers to all questions regarding the effects of delays, for every possible type of system or task. Neither does there seem to exist a complete theoretical or practical framework for the management of system delays in, for example, software development (cf. Dabrowski & Munson, 2011; Chen et al., 2007).

It does seem likely, however, that sensitivity to system delay is largely dependent on the type of task performed, as well as where and when delays occur in said task. Furthermore, continuous systems seem to be more sensitive to delays than discretionary systems, because of the more tightly coupled closed-loop control. There also seems to exist a positive correlation between the relative size of the controlled artefact – in the sense of computer screens, or virtual reality goggles – and sensitivity to system delays (c.f. Beigbeder et al., 2004; Ellis et al., 2004). Things that cover more of a user’s vision generally have higher demands when it comes to delays. Generally, though, the adverse effects of delays seem to be well known within the field of continuous systems.

For discretionary systems, however, things are not as straightforward. Some delays have been shown to improve user performance, and is thought to be because they do not disrupt the natural work-flow of the user. These so called inter-task delays seem to give users more time to think and reset before moving on in their work, thus improving their subsequent interactions. Inter-task delays should be of a medium length, and what constitutes as medium is, again, dependant on the task (Hoxmeier & DiCesare, 2000; Selvidge et al., 2002; Sellier & Chattopadhyay, 2009, Kohlisch & Kuhmann, 1997). Task interruptions, however, have unanimously been shown to have detrimental effects for user performance and satisfaction. The idea that there exists two types of delays – interruptive and inter-task delays – can be condensed in coined the ‘theory of task interruption’. This theory would imply the nature of delays, in that they can be of either an interruptive or of a non-interruptive type, and will henceforth be used in describing this distinction.

The article by Teal and Rudnicky (1992) – studying choice and change of user strategies – as well as many of the following replications (e.g. O’Donnell & Draper, 1996) found that system delays have an effect on how the users interact with a system. Not only do system delays seem to affect user performance in some fashion, but they also seem to have an effect on how a task is approached and completed. This is most strikingly evident in the example of space exploration, where the operators controlling far-away space-craft have opted to use a wait-and-see strategy in their control, instead of trying to control them continuously (see Lester & Thronson, 2011).
Additionally, a few recent studies have introduced concepts such as predictability and predictivity of delays. These aspects can be considered valuable to systems, in that they can both be used as feedback, or otherwise pieces of information in the design of a system. The theory of task interruption, as well as these two examples of how delays can be used in creative ways in design, goes to show that system delays are much more than just a simple computer response time.

2.3 Outlining the experiment

Given the current understanding of the possible effects of delays in discretionary systems, and the apparent lack of studies conducted on sub-second delays, it is of interest to examine these aspects of interaction. Furthermore, sub-second delays are a prevalent tribulation in many of today’s systems (Szameitat et al., 2009), and will most likely continue to be so for some time. An experiment was therefore conducted to test the effects of sub-second delays for discretionary systems, in the form of simple computer interaction. To simulate this type of interaction, and its underlying cognitive processes, the Trail Making Test (TMT) was chosen. It is hypothesized that the TMT is complex enough, and that the users are under sufficient cognitive load as to be significantly affected by different system delays. As the TMT requires use of visual scanning, motor speed and attentional abilities (Wikman & Summala, 2005), any effects of system delays will have an observable effect on the performance of the users.

Contemporary research seems to agree that the important factor in determining if a delay will have detrimental effects on users is where in a sequence of actions it is presented (e.g. Hoxmeier & DiCesare, 2000; Selvidge et al., 2002; Sellier & Chattopadhyay, 2009, Kohlisch & Kuhmann, 1997). If a delay constitutes as a task interruption it will have a negative effect on user performance and satisfaction. Delays in the TMT are hypothesized to be of an interruptive nature, as delays will hinder the users from entering the sequence of numbers and/or letters (see section 3.3 for detailed description of TMT) as quick as possible. It is therefore hypothesized that the results will show a negative correlation between user performance (response times and error-rate) and system delay. It should, however, be noted that if the delay does not disrupt the users, and if the users can maintain a constant work-flow regardless, then delays should not produce any negative effect on the users’ performance (see Kohlisch & Kuhmann, 1997; Shneiderman, 1984; Barber & Lucas, 1983).
3 Method

3.1 Participants

Participants (N=30, 16 F, 14 M) were students at Linköping University and were between the ages of 19 to 30 (Mean=22.7, SD=2.3) years old. For their participation they received a cinema voucher (approx. value 110 SEK). Participants were naïve as to the purpose of the experiment, and were only told they would be performing a simple form of computer interaction, so to avoid raising awareness of the delays. Participants were required to have normal (or corrected to normal) eye-sight, as well as not having any impairments that could affect the interaction with a computer (such as neurological disorders or physical disabilities).

Before the test, participants were informed of their rights; that the test was completely voluntary and anonymous and that the collected data would be handled confidentially. A consent form was then signed. The experiment was conducted in accordance with the ethical guidelines of the Declaration of Helsinki (2008).

3.2 Apparatus

![Figure 3: The author, illustrating the interaction with the screen](image)

The experiment was run on an ordinary PC laptop, connected to a 23” touch-screen monitor. The participants were seated at about arms length (approx. 40 cm) from the monitor, and were free to use either (or both) of their hands in the interaction with the monitor. All interaction was done by touching
the screen (see Figure 3). The room in which the test was conducted was quiet, well lit, and had little in the form of possible distractions.

3.3 Procedure

The experiment buildt upon the digitalised Trail Making Test (TMT) developed by Summala et al. (2008), in turn, based on the pen and paper version developed by Reitan (1958). TMT originally required the participant to draw lines between only numbers (variant A) or numbers and letters combined (variant B) in alphabetical and/or numerical order (see Bowie & Harvey, 2006). The digital TMT, however, makes use of a touch-screen, where participants press the numbers and letters on the screen (see Figure 3), in succession, instead of drawing a line between them, as they would in the original pen and paper version. Figure 4 illustrates the difference between the A and B variants of the digital TMT.

In addition to the A and B variants, the digital TMT adds two other conditions to the experimental situation, namely fixed or random sub-variants of the A and B tests (henceforth subtasks). The random variant of the A and B tests change the position of the numbers and/or letters, so that participants have to scan the screen for the next letter/number for every time they advance in the numerical/alphabetical order (i.e. press a letter/number on screen). The circles themselves do not change position, only the number/letter within them. The fixed variant works more like the original TMT, in the sense that numbers and letters stay in place, and the participant is free to press these at his/her own leisure.

![Figure 4: Illustrating the A and B variants of the TMT](image)

A slight modification was made to the digital TMT (i.e. Summala et al., 2008) involving the addition of delays in between actions, consisting of a white (blank) screen shown after the participant had engaged with the touch-
screen. The experiment consists of four experimental conditions, one for each delay condition, namely 0, 250, 500 and 750 msec. The 0 msec delay condition is essentially the digital TMT in its original form, without any added delay. Upon touching the screen a small ripple effect was displayed as to indicate a touch had been registered. The experiment used a within group design, where the different experimental conditions (i.e. delay conditions) were presented in random order for each participant, as to minimize any effects of primacy or fatigue.

The experiment was initiated with a short introduction to what the participants would be doing. This was followed by the reading and signing of the consent form. Included was also an information sheet, explaining the TMT test in its entirety. When the participant felt ready, the experiment leader initiated the test. Once the computer program had started the participant was left to follow the instructions on screen (which were present between every sub-task in the test). Once the first test (i.e. a randomly selected delay condition) was completed, the experiment leader manually initiated the next one, in the previously determined random order of the four experimental conditions. Once all the tests had been completed the participant was rewarded with a cinema voucher and the experiment leader offered to answer any questions the participants might have. In its entirety, the experiment took about 45 minutes to complete.

<table>
<thead>
<tr>
<th>Order</th>
<th>Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A Fixed</td>
</tr>
<tr>
<td>2</td>
<td>A Random</td>
</tr>
<tr>
<td>3</td>
<td>B Fixed</td>
</tr>
<tr>
<td>4</td>
<td>B Random</td>
</tr>
<tr>
<td>5</td>
<td>B Random</td>
</tr>
<tr>
<td>6</td>
<td>B Fixed</td>
</tr>
<tr>
<td>7</td>
<td>A Random</td>
</tr>
<tr>
<td>8</td>
<td>A Fixed</td>
</tr>
</tbody>
</table>

Table 1: Shows the order of which the different subtasks were performed

Each individual test (delay condition) started with four trials of the different test variants (A/B, Fixed and Random). After these four tests were completed, the real part of the test (i.e. the eight subtasks illustrated in Table 1) was initiated. Each subtask was conducted twice, in reverse order, and that response times were measured and averaged for both subtasks.
4 Results

4.1 Response time

The results were analysed for outliers, leaving 29 participants for further analysis. Outliers are defined as individual reaction times that exceed two and a half standard deviations from the mean of a specific subtask. Only two outliers were found (after removal of the one participant) and were changed to the next highest score plus one ($N_{MAX} + 1$). All outlier reaction times (including those of the removed participant) were on the positive end of the spectrum, which is indicative that some distraction caused the reaction time to become prolonged, for these specific subtasks. After removal of outliers, the data was checked for normality using the Kolmogorov-Smirnov test. The reaction time data was normally distributed, with the exception of the subtask of A Random, for delay conditions of 250 and 750 msec. Data was analysed using a repeated measures ANOVA, with the exception for A Random, that instead was analysed with the non-parametric test of Friedman. Within the ANOVA, a Bonferroni correction was used so to reduce the risk of Type 1 errors. Error bars show ±1 standard deviation.

4.1.1 Subtask A Fixed

Figure 5 shows the group mean reaction times for the four different delay conditions for the subtask A Fixed. The mean reaction times presented have been normalised for the delay itself. The results of the repeated measures
ANOVA were not significant (F(3, 84) = 1.422, p = .242) (Mauchly’s test n.s.).

4.1.2 Subtask A Random

Figure 6 shows the group mean reaction times for the four different delay conditions for the subtask A Random. The mean reaction times presented have been normalised for the delay itself. The results of the Friedman test were not significant ($\chi^2 = 3.538, p = .316$).

Figure 6: Mean reaction times for in msec subtask A Random

Figure 7: Mean reaction times in msec for subtask B Fixed
4.1.3 Subtask B Fixed

Figure 7 shows the group mean reaction times for the four different delay conditions for the subtask B Fixed. The mean reaction times presented have been normalised for the delay itself. The results of the repeated measures ANOVA were significant, $F(3, 84) = 2.948, p = .037$ (Mauchly’s test n.s.).

![Mean Reaction Time](image)

Figure 8: Mean reaction times in msec for subtask B Random

4.1.4 Subtask B Random

Figure 8 shows the group mean reaction times for the four different delay conditions for the subtask B Random. The mean reaction times presented have been normalised for the delay itself. The results of the repeated measures ANOVA were not significant ($F(3, 84) = 0.726, p = .539$) (Mauchly’s test n.s.).
4.2 Error-rate

Results are reported for 30 participants. Only three outliers were found (being defined as a value larger than 2.5 standard deviations from the mean) and were changed to the next highest score plus one ($N_{MAX} + 1$). Data was thereafter checked for normality using the Kolmogorov-Smirnov test, which showed significance for all four conditions. Friedman’s test (which does not assume normal distribution) was therefore used to test for effects.

Figure 9 shows the group total error-rate for the four different delay conditions across all four subtasks. The results of Friedman’s test were not significant ($\chi^2 = 1.925$, $p = .382$).
5 Discussion

5.1 Discussion of experiment

The experiment yielded but one significant result, a positive performance increase on subtask B Fixed given longer delays, with the other three subtasks seemingly unaffected by the delay. The assumed hypothesis that delays would result in a negative effect on performance could not be proven. In fact, the results do not show any such tendencies.

The subtask that did show a significant linear effect (B Fixed, where numbers and letters are combined in a stable environment), and to some degree the other three subtasks, indicate that the chosen delays were not of an interruptive nature, but were rather inter-task delays, natural pauses in between interactions with the screen (see Hoxmeier & DiCesare, 2000; Selvidge et al., 2002; Sellier & Chattopadhyay, 2009; Kohlisch & Kuhmann, 1997). The results support the theory of task interruption, and further that inter-task delays can have positive effects on users (given a moderate delay), as the users get more time in between each interaction with the screen. The performance increase in user response time, for subtask B Fixed, from the no-delay condition to 750 msec, was only about 12%. More surprising perhaps, was the general improvement in error-rate across all subtasks, with users making about half as many errors at 500 and 750 msec delay compared to the no delay condition. The error rate was not significant, and it should only be seen an indicative of any underlying effect, as its validity can be questioned. Some users may have been more error-prone than others in the no-delay condition, something that was indicated by skewed data.

The measured performance increase in conditions with longer delays – specifically for subtask B Fixed – could potentially be due to the simple fact that users had more time to think (see Teal & Rudnicky, 1992), in that part of their response time on the no delay condition was taken up by “think time”, and that giving users a delay shortens their response time by an equal amount. Were this the case however, the positive effects of adding delay should have been proportional to the length of delay, when in fact the performance increase is only a fraction of the time added by the delays. This indicates that the users do not use the delays as additive think time per se, but rather use the extra available time to do something else. The error-rate does suggest that users adopt a different strategy in the interaction as delays are increased (see Teal & Rudnicky, 1992), while still managing to respond equally fast in the no-delay condition for most of the subtasks.

If the measured effects of delays are because users in this case benefit from longer, rather than shorter, delays (in accordance with Kohlisch & Kuhmann, 1997; Shneiderman, 1984; Barber & Lucas, 1983), then why does
only one of the subtasks show this effect? Granted, the three subtasks that did not show any significant positive effect of delays are not showing any negative effect of those delays. Still, it is unclear why only the subtask of B Fixed shows an improvement in response time given longer delays. It seems strange that tasks so similar could be affected differently by delays. A possible explanation for this discrepancy is that the B Fixed was specifically susceptible to improvement by a slower interaction pace, where the others were not. After all, the improvement in user performance was not substantial.

With regard to the failure to confirm the hypothesis, Dabrowski and Munsen (2011) explain the lack of measurable negative effects of longer delays by "if the user can keep his or her train of thought or can keep making progress toward a larger goal, total task completion times can be unaffected even by large delays" (p. 559). This is indeed also in line with the theory of task interruption. The authors advocate that inter-task delays have positive effects for users, and this study seems to give further evidence for this theory.

5.1.1 Ecological validity

All experimental research should concern itself with the possibility of not being representative of a naturalistic setting – this study being no exception. The experiment was conducted in a laboratory setting, artificially testing users and their cognitive abilities, in the hope of simulating the workload of a complex and challenging computer task. However, there are objections to laboratory studies being representative of a naturalistic setting, for example those presented in Klein et al. (2003). The authors argue that cognitive abilities function very differently in a naturalistic setting (macrocognition), compared to a laboratory settings (microcognition). They argue that cognitive abilities work very differently because of a larger complexity, users feeling pressured, higher risks being involved, as well as ill defined goals present in such a setting. Be that as it may, the effects of delays should, in the case of the conducted experiment, be more prevalent given a naturalistic setting, due to the higher demands of such tasks. The results of previous studies in the field, as well as the current one, stand as indications to the effects of delays in a naturalistic setting.

Of course, the best possible scenario would be to test the effect of delay in an actual real world situation (e.g. Barber & Lucas, 1983). Such an undertaking would, however, be many times more difficult than a laboratory study and further unreasonably time consuming. Besides, a naturalistic setting would not allow for the amount of control over the different experimental conditions available in a controlled laboratory setting.
The digital TMT in the conducted experiment is further seen as translatable to a real world task, as it tests the underlying abilities used in such tasks, more so than it is a completely artificial and arbitrary challenge. It does not simply create a meaningless task for the users to perform, but challenges their cognitive abilities in areas commonly used in real world tasks. User performance and interaction should therefore be affected equally in the TMT, as it would in a real world situation, given different delays.

5.1.2 Limitations

One limitation of the conducted experiment is the absence of satisfactory feedback to the user in the interaction. The digital TMT was originally designed to give users feedback in form of a clear sound, for every interaction with the screen performed. This was not present in the conducted experiment. Instead, the screen displayed a [barely visible] ripple effect on the screen when the user interacted with it. The absence of feedback to the users could potentially have had significant effects on the results – it could be the reason why some participants made many more errors in the no delay condition compared to the other conditions. In the no-delay condition, there really was no clear indication that the computer program had registered an interaction. This was especially prominent in situations where some of the other delay conditions were performed before the no-delay condition (where the delays, i.e. white screens, acted as natural feedback and proof of successful interaction). Several participants expressed confusion over the lack of feedback on the no delay condition during the experiment.

Given the importance of instant feedback of action initiation (see section 2.1.1, Figure 1; Johnson, 2010, p. 160), this is potentially a serious problem for the validity of the results. This omission, however, was not due to oversight or ignorance, but because of the technical limitations of the present equipment. Any future replication of this type of interaction should keep in mind the importance of satisfactory feedback.

Another limitation is the choice to study only sub-second delays. It would have been interesting to examine the effects of delays beyond 750 msec on the TMT, and what their effects would be. Again, this was limited by the experimental situation; every new condition would have added another 10 minutes to the duration of the experiment. Given the difficulty of finding 30 participants for a 40 minute test, even with the reward of a cinema voucher, it was a trade-off decision between acquiring more data versus actually getting participants. Participant fatigue also becomes a problem when experiments linger on for too long.
5.2 General discussion

The aim of the current study was to investigate temporality in interaction, which proved to be anything but a trivial subject area. The background has both defined a certain perspective on what temporality in interaction means, following the paradigms of human factors and HCI, as well as consolidated a number of previous publications in the field. The complexity of the issue becomes apparent when delays transform from being just system response time to something entirely different.

The conducted experiment examined constant, sub-second delays, and their effects on user performance. The results support some of the contemporary research on the area, in that some types of delays have positive effects for users (Kohlisch & Kuhmann, 1997; Shneiderman, 1984; Barber & Lucas, 1983). However, as is evident from the review of literature, there are many more aspects to temporality other than those examined. Other than constant delays – which was the focus of the current experiment – there are variable delays, which have been shown to be especially detrimental to user performance (Thomaschke et al., 2011; Thomaschke & Dreisbach, 2013; Rolke & Hofmann, 2007), as well as delays beyond a second in length. The current study opted to examine delays for a simple discretionary system, whereas there are also continuous systems, as well as different categories of systems within those two. There are also many aspects of delays that go beyond the simple “wait-time”. Predictivity and predictability are areas of system delays that are novel concepts within the field, offering a whole new perspective on system delays (Thomaschke & Haering, 2014; Wagener & Hofmann, 2010) and can grow to become an integral part of designing for (or around) system delays. Temporality in interaction has proven to be as complex as any other part of system design. System delays are more than just a system response time, in that they, in themselves, can be informative and part of a system’s structure.

5.2.1 Theoretical implications

The length of constant system delays and their effects on users have been the major point of scientific inquiry in the past half century, which has resulted in a well documented understanding of the effects of such delays (e.g. Szameitat et al., 2009; Hoxmeier & DiCesare, 2000; Galleta et al., 2006; Dabrowski & Munson, 2011). However, there are many more aspects of delays than just their length.

Two such aspects of the field are predictivity and predictability, that were until only recently unheard of within the field. To reiterate the background (see section 2.2.2); predictivity is the delays’ ability to in themselves act as
feedback, in that a certain delay length, deterministically or probabilistically, implies a specific system output. Predictability of delays is the ability for users to predict the length of a delay based on constancy of delay length. Only a few studies have examined the effects of predictivity and predictability (see Thomaschke & Haering, 2014; Wagener & Hofmann, 2010), and it constitutes an area where more research is needed, as to fully understand the effects of these aspects. Questions as to what effects predictivity can have on user performance (cf. Thomaschke & Haering, 2014), what the positive effects of predictable delays are (if they, in fact, can mitigate the negative effects of the delays themselves, cf. Thomaschke et al., 2011), are two of many remaining questions with regard to these aspects of system design.

Variability of delays is further a related concept to the predictivity and predictability of delays. Variable delays, unlike their constant equivalents, are inherently unpredictable in stochastic systems. Variable delays require external feedback (in the form of progress-bars) to become predictable, as the users cannot acclimate to constancy in delay length. Variability of delays and predictivity, however, are not mutually exclusive, but rather two sides of the same coin. The term variability is associated with, and governs the (often) unpredictable and stochastic nature of variable delays. Predictivity implies causality between delay length and system response, often a deterministic one. For predictivity to be effective there must exist more than one type of system response, and likewise a specific length of delay for each of those conditions. The negative effects of variability have seen some scientific exposure, but would benefit from more research, as the severity of variable delays is not fully established (cf. Thomaschke et al., 2011; Thomaschke & Dreisbach, 2013; Rolke & Hofmann, 2007).

Another seeming gap in the literature is found in the study of task interruptions and inter-task delays. Convincing, albeit not indisputable, the theory of task interruption currently stands strong. The results of the conducted experiment in the current study seems to support the existence of inter-task delays – that is, natural breaks in work-flow where delays can occur and not impede user performance. This theory has little in the form of explicit support, but many previous publications speak in its favour (e.g. Kohlisch & Kuhmann, 1997; Shneiderman, 1984; Barber & Lucas, 1983; Dabrowski & Munson, 2011; Johnson, 2010).

These three aspects of discretionary systems – predictivity and predictability, variability of delays, and the theory of task interruption – can grow to become prevalent components, or even the main focus, of future research in the field. They have the potential to change how we see system delays, and how we can design for, or around, delays where they pose a problem.

Beyond the tribulations of system delays found in discretionary systems, there also exists the category of continuous systems, and its own problems
therein. Delays in a continuous system appear to only be of an interruptive nature, regardless of the nature of the task or its complexity. The threshold for users’ perception and disruption of delays in such a system appears linear to the accuracy needed in the completion of the task (e.g. Lester & Thronson, 2011; Chen et al., 2007; Pantel & Wolf, 2002; Allison et al., 2001). So far as the effects of delays in continuous systems are concerned, the current understanding is ample. An area that could benefit from more academic inquiry is that of the possible solutions to unavoidable system delays, for example the anticipatory feedback in the form of probability based screen rendering presented in Chen et al. (2007). Another area of interest is perhaps the one of strategy selection in the control of continuous systems, when a user transitions from exerting continuous control to a wait-and-see strategy – at what point does this transition occur, and what are its effects, are a few, among many, questions still unanswered.

Furthermore, there is cause for concern with regard to the lack of encompassing theories of temporality. Given the stark difference between discretionary systems on the one hand, and continuous systems on the other, it would seem beneficial to adopt such a distinction in the pursuit of more general, and particularly applicable, theories on the effects of delays in interactive systems. Dabrowski and Munson (2011) made a similar distinction, but chose to call the two types of system conversational tasks and control tasks, respectively. This sort of distinction – whether the one presented in the current study, or that of Dabrowski and Munson, is chosen – is necessary before any truly major theoretical assertions can be of any value (ibid, 2011). More studies with a holistic perspective are needed to fully understand the ramifications that govern the effects of delays on users.

5.2.2 Practical implications

Practical implications are one of the more important aspects of research, as knowledge and understanding does little good when hidden from actual worldly applications. The field of scientific inquiry should together with an innovative industry drive technology and practical solutions forward. So what can be said about temporality in interaction, and especially the practical implications of our understanding of system delays in discretionary systems, given the contemporary research?

As per the theory of task interruption, system delays should not be regarded as an absolute detriment to user performance. Certain types of systems (or tasks within them) allow for natural breaks in user work-flow, hence allowing for some delay between interactions (i.e. inter-task delays, see Hoxmeier & DiCesare, 2000; Selvidge et al, 2002; Sellier & Chattopadhyay, 2009, Kohlisch & Kuhmann, 1997; Johnson, 2010). This delay needs to be of a
medium length for this effect to remain true – as excessively short or long delays can be detrimental to user performance (Kohlisch & Kuhmann, 1997). What constitutes as a medium delay is entirely dependent on the type of task being performed. Even though there exists no explicit guidelines (yet) as to how to evaluate whether a delay constitutes an inter-task delay or a task interruption, it would be wise to keep in mind that there does, in fact, exist such a distinction. Delays should, therefore, not be avoided in principle, but because they are believed to interrupt users.

Some studies have further shown that some inter-task delays can result in a performance increase, beyond not being interruptive to users (Kohlisch & Kuhmann, 1997; Shneiderman, 1984; Barber & Lucas, 1983; Dabrowski & Munson, 2011). Delays can thus be used to help users maintain a sensible interaction pace. Studies have shown that too short a delay can cause users to change to a more inaccurate, albeit faster, interaction strategy (Teal & Rudnicky, 1992; O’Donnell & Draper, 1996). Designers have the power to influence user interaction strategy by varying only the duration of the delay; a shorter delay allowing users to interact faster, but potentially at the cost of lower accuracy/higher error rate. The longer a user has to wait to input a set of instructions/data, the higher the cost of making errors and, likewise, the more accurate the user will be in interaction (Teal & Rudnicky, 1992).

Furthermore, delays can be used as pieces of information, as an integral part of the feedback framework within a system in the form of predictivity. By controlling the delay length based on the expected system response, the user can remain informed of system status, without engaging in anything other than just ”waiting” (see Thomaschke & Hearing, 2014). A good implementation example of this is a system which always gives an error message within a set amount of time, and always delays longer when the users has entered a set of instructions correctly. If the set amount of time has passed, and the system has not given an error message, the user can rest assured that the system is processing his/her information, and assuming nothing else goes wrong, know that the command will be executed. This is feedback, as it informs an experienced user of the status of the system, potentially as much as any dialogue box would have. Moreover, a system can employ the use of predictability in delays (something many of today’s systems already do in the form of progress-bars or animated indicators, see Branghan & Sanchez, 2009), but also merely by constancy of delay length. Predictability can allow users to maintain a stable interaction pace, in that they are suspending their interaction trough an approximation of the remaining delay, instead of on empty waiting time.

Both predictivity and predictability require a rather omniscient system – a system that knows what all its possible outputs will be. This might be counter-productive, if a requested process is novel to the system, it will have
no way of knowing how much time is actually remaining on the process – effectively, how long the delay will last. Often, however, this is not the case; a system often processes the same types of information countless times, and for extended periods of time. Implementing a semi-deterministic (probability based) predictive or predictable system seems feasible for most system types, and something designers can do to alleviate the (potential) negative effects of prolonged system delays. It would be wise to remember, however, that too much feedback is not always optimal either (see Branghan & Sanchez, 2009).

In the case of variable delays, which amounts to an especially detrimental type of delay, they can be alleviated by external predictability. Even if delays are of varying length, a progress bar, for example, may provide users with sufficient feedback for them to maintain a stable and consistent interaction pace. However, a few studies (e.g. Sellier & Chattopadhyay, 2009) have suggested adding artificial delay to especially short delays, as to homogenise an existing variance in delay lengths, so as to achieve constancy of delay length. This type of solution would add artificial time to short delays, thus creating a more even distribution of delay lengths (e.g. Szameitat et al., 2009). Another way of minimizing variability of delays is scheduling, i.e. the distribution of computational processes over a longer period of time (see Thomaschke & Hearing, 2014).

With regard to continuous systems, the contemporary research stands clear: delays should be minimized, or entirely avoided if at all possible (e.g. Lester & Thronson, 2011; Chen et al., 2007; Pantel & Wolf, 2002; Allison et al., 2001). The very nature of a continuous system is that control is exerted instantaneously by the user, and when the system does not respond in a timely manner (going beyond the threshold for a certain type of system) the user often has to change to another type of control strategy. Some continuous systems are inherently unable to be controlled without using a wait-and-see strategy (due to being at a distance yielding a few seconds transmission delay, see Lester & Thronson, 2011). There are, however, proposed alleviations to this particular problem, in the form of predicting how a system would respond given a certain input (see Chen et al., 2007).

5.3 Are system delays soon a thing of the past?

A reasonable assertion of today’s technologically driven world is that our systems are becoming faster and faster, that their computational power is increasing with every year, ad infinitum. Given that system delays are mostly due to the insufficiency of computational power (an exception, of course, being systems limited by transmission delays), it seems likely that, sooner or later, the available technological resources will outgrow the demand, hence
rendering research on delays obsolete, and system delays a thing of the past. To digress; is this a desirable development that will contribute to better systems?

Continuous systems are unanimously sensitive to delays, as the negative effects on users are substantial. However, discretionary systems do not tell the same simple story, but represent a more varied spectrum of user requirements. The fact that some system delays might benefit users may have come as a surprise to many, as delays have long been revered as a technological disease, and a symptom of slow hardware; our systems destined to be completely void of such problems in the future. Given the potential positive effects of delays (and the negative effects, to the contrary, in the absence of delays) working towards minimising and removing all delays across all types of systems would be unwise. System delays should be minimized because of their interruptive nature, and kept if they benefit the users.

The acclaimed benefits of delays notwithstanding, some systems may profit from delays, not because they entail some potential performance increase to users, but because they have additional virtues. To illustrate, imagine a system that takes some words as input from the user, and returns a series of relevant articles based on that input (a search engine, basically). A prospective user wants to find a few relevant articles in a multi-million item database, and enters a few common terms on the area of interest. The system then proceeds to, without delay, present a few thousand search hits. However, because the user understands the magnitude of the requested action, the presented results seem questionable; it should have taken at least a few seconds to search such a database. Thus, the user questions whether the system has actually performed the requested action, or simply produced a random output. This example illustrates merely one area where the lack of a delay affects the users’ perception of a system; in this case the reliability. Sellier and Chattopadhyay (2009) proposed the adding of delays to systems where too short delays would indicate that "something was not right". Performing a complex action should take some time, as users will otherwise question whether the system did anything at all. This effect can be seen as a form of reversed predictivity; the user imposing his/her own mental model of what should be a reasonable delay length before a certain system output. Designers should take heed of users’ presumptions for all types of systems, in that the system should match the expectations of the user, or else it will seem less reliable.

Taken together, the systems of tomorrow may still have delays – where applicable – not because the hardware arbitrarily limits the system, but because the designer chose a constant and consistent, deterministically predictable, non-interruptive delay, in the design of the system, so that user performance and satisfaction could be maintained.
6 Conclusions

As is evident from the many previous studies conducted in the field, temporality represents an integral part of interaction. A categorisation can be made between discretionary systems on the one hand, and continuous systems, on the other. This division makes sense in that the temporal limitations of users are similar to either one of the two types of systems defined. Discretionary systems abide by no specific temporal guidelines, but rather of the type of task being performed in said system. Delays can therein be either interruptive, or non-interruptive (inter-task delay). Non-interruptive delays are found in tasks that allow for natural pauses in work-flow, and can even potentially improve user performance and satisfaction in those systems. Furthermore, delays can be either informative or non-informative, in that they can be used as, or instead of, feedback in a system. Continuous systems, however, are unanimously affected by delays to their detriment, and should therefore be completely void of delays for best user performance and satisfaction.

The current study has examined constant sub-second delays for an elementary (but still cognitively demanding) discretionary task, and found that the presented delays were not of an interruptive nature. For one of the subtasks in the experiment, a significant improvement in user performance was measured when delays were increased. This result stands to support the presently defined theory of task interruption; in that some delays, if presented in a natural pause in interaction, can have positive effects on user performance. The general results also support the concept of delays not being detrimental to users when presented in specific types of tasks.

Humans are not machines, but limited, and indeed defined, by our innate abilities. Interactive systems should abide by those, and be designed for the user and her virtues. Systems should even consider the expectations of the user, and maintain and support them with appropriate delays; an area where artificial delays could show their potential worth. The current study stands clear on the point that the user should always stand at the centre of the interaction. Furthermore, the ideas and theories presented herein may come as a surprise to many, that delays are not unanimously detrimental to users regardless of situation, and further that there is a depth to the issue of temporality in interaction, beyond the simple "wait-time".

In due time, when all interactive systems have enough computational power to be rid of delays, it should be remembered that only the interruptive delays should be removed. The informative and otherwise favourable delays should remain, and so help ensure a user friendly and effective interactive system.
7 Future research

Much still remains to be studied when it comes to the field of temporality, as the true nature of the human mind remains not fully understood. The subject of temporality in interaction has proven to be an area as complex as any when it comes to human interaction, and will most likely continue to vex scientists and designers alike in the coming decades (see section 5.2.1 for concrete examples of possible research questions).

Future research is likely to involve the positive and potentially useful sides of delays, more so than the already rather well documented detrimental effects. These positive, and indeed applicable, aspects of delays can be of great value in the continuing development of better and smarter interactive systems. Knowing how to design around and for delays, rather than against them, will in all probability be of more value in the coming years. New and innovative solutions to the problems caused by delays, as well as improvements by delays themselves, are presumably to be the main focus of the academic world as well as the industry in the near future.

Likewise, future research can concern itself with the possibility of delays being completely eliminated within the next decade. What would the effects of systems completely devoid of delays be to users? This question poses a valid concern for the performance and well-being of users in such a potential world.

Furthermore, the current study has presented a few possible alleviations to arising issues of temporality, but the fact remains that no technological development ever stands complete. The systems of tomorrow, whatever their interaction may be, will have their own unique set of problems regarding temporality. There is no doubt that both the scientific community and industry will concern itself with these new, and potentially exciting, fields.
References


Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a considerable time from the date of publication barring exceptional circumstances.

The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: http://www.ep.liu.se/

© Martin Krampell