The Influence of Cognitive Abilities and Cognitive Load on Business Process Models and their Creation

Manuel Neurauter, Jakob Pinggera, Markus Martini, Andrea Burattin, Marco Furtner, Pierre Sachse, Barbara Weber

University of Innsbruck, Austria
{manuel.neurauter,jakob.pinggera,markus.martini,andrea.burattin, marco.furtner,pierre.sachse,barbara.weber}@uibk.ac.at

Abstract. While factors impacting process model comprehension are relatively well understood by now, little is know about process model creation and factors impacting process model quality. This paper proposes a research model to investigate the influence of cognitive abilities and a continuous psycho-physiological measure of task imposed cognitive load of process model designers on process model quality. The proposed research will not only contribute a better understanding of process model creation, but bears significant potential for improving existing modeling notations as well as for developing process modeling environments.

Keywords: cognitive load · working memory · executive functions · reasoning ability · business process modeling

1 Introduction

According to [1] conceptual models are used by practitioners for analyzing business domains and for an easier development of information systems. Relevant information regarding the business domain, such as states, events, tasks and business rules are illustrated in various graphical and textual notations as business process models [2]. These process models often play a crucial role in re-designing business domains and in organizational analysis [3], and due to the wide range of problems displayed by industrial process models [4,5] an in depth understanding of factors influencing process model quality is in demand. Past research has shown that complex process models tend to contain more errors [4], whereas modeling expertise [6], [7], process knowledge [7], activity labeling [8], routing symbol design [9], notational deficiencies [10], and *cognitive abilities*, learning style and learning strategy [11] provide measurable impact on process model comprehension. Moreover, it has been demonstrated that characteristics of the modeling task are influencing process model quality [12]. While factors determining process model comprehension are relatively well understood by now, only a few studies focused on process model creation (e.g., [4], [13, 14]).

The creation of process models is characterized as design activity [15,16] and imposes a variety of challenges which include the construction of a mental model of the domain as well as the externalization of the mental model by mapping the mental model to the modeling elements provided by the modeling notation using a modeling tool [17]. The cognitive demands imposed on the process model designer (designer for short) herby depend on task-specific factors like the task's inherent complexity, the modeling notation, and the modeling tool support. These demands are commonly described as *cognitive load* (CL) [18,19]. In addition, the crucial role of *cognitive abilities* for process model quality is stressed, e.g., [10], [15], [20]. To gain deeper insights regarding the creation of process models cognitive abilities as well as the occurring CL should be considered (e.g., [10], [20].

Our research will be a similar approach as [10], but shift the focus from process model comprehension to process model creation. This work will provide a better understanding of factors impacting process model quality by integrating task-specific factors through continuous CL measurement and human factors in form of cognitive abilities into a single study. In the remainder of this paper we will provide theoretical backgrounds and introduce the research model including research questions.

2 Theoretical Background

Following [15,16], and in line with [14] we interpret process model creation as a cognitive *design activity* within the field of problem solving [13]. As pointed out previously, our research will focus the cognitive abilities of designers and the influence of task imposed CL on process model quality. As stressed by [20] cognitive abilities such as *reasoning ability* (RA), *working memory* (WM) and *executive functions* (EF) are crucial for creating process models of high quality, and for design activities, and problem solving in general [13]. In addition to the cognitive abilities the designer's CL plays a key role for problem solving and design activities [21,22,23].

To assess the interaction of those cognitive abilities and task specific factors within a long-running design activity, we are going to use psychophysiological measures of CL.WM, RA, and EF therefore are our independent constructs, and the quality of process models stands as dependent construct. CL takes a special role in our research, because of the possibility to utilize CL either as independent construct or as dependent construct, as pointed out in the following section.

CL, mental effort, mental load, and mental workload are widely used as aliases, basically describing the same concept [18]. According to [19], CL characterizes the demands of tasks imposed on the limited information processing capacity of the brain in the same way that physical workload characterizes the energy demands upon muscles. CL therefore represents an individual measure considering the individual amount of available resources and task-specific factors imposing CL. As independent construct, CL predicts performance for task execution, since high CL leads to poor task-performance and to wrong decisions, e.g., [21,22,23]. CL, on the other hand, is influenced by cognitive abilities. As pointed out by [21], RA reduces load on WM by utilizing former knowledge and experience by linking strategies to goals and both,

WM and RA are guided by EF as they regulate thought and action, e.g., [20,21], [24]. This leads to CL as dependent construct. During process model creation the designer will face a variety of task-specific challenges like the construction of a mental model of the domain and the externalization of the mental into a process model. This requires the usage of different modeling elements (e.g., gateways, activities, edges) and chunks of these modeling elements, e.g., when creating loops. While existing studies on process model comprehension typically assess cognitive load once after task completion [10], [25], this is not sufficient for accurately assessing the cognitive demands implied by a (long-running) design activity. However, applying a continuous measurement of CL with high temporal resolution such as pupillometry and heart rate variability [21], additionally to the post-hoc assessment by a widely used questionnaire (i.e., NASA-TLX [26]), allows to investigate CL regarding both, process model quality and task specific factors. For this, the measurements of CL either can be aggregated for the entire task with the overall quality in scope, or for task specific factors of interest.

RA refers to the process of drawing conclusions or inferences from different information. This always requires going beyond the information that is given and, thus, is closely related to other domains of human intelligence [24]. The reasoning process within process modeling can be described as combining given environmental input about a domain with previously made experiences or knowledge with the objective of externalizing a model representing the actual domain [20]. Previous knowledge and experience within a domain therefore enhances the reasoning process and helps to link strategies to goals freeing WM resources for problem solving tasks [27,28,29]. To assess RA, psychometric tests like, e.g., the raven's progressive matrices [30] will be utilized.

The cognitive system of WM provides temporary maintenance of relevant information required for task performance [31,32]. Because of its limitation to about four items [32] WM is a central predictor for inter-individual differences in complex cognitive tasks, e.g., [33,34,35] including process model creation [13]. Therefore, WM is mostly defined as a construct consisting of a set of cognitive processes, with at least two distinguishable components, namely *holding and processing* and *relational integration* [36,37]. Holding and processing quantifies the ability to hold limited amounts of information (e.g. letters, symbols) outside the focus of attention, while other information (e.g. calculations, sentences) is processed simultaneously. Relational integration measures the ability of building new relations between elements such as single dots into a pattern [38]. The central role of WM in process modeling is well known [20], [25], [39,40,41], but most often only theoretically implemented. Only a single study empirically tested the role of WM for process model creation [13] and one study focused on related concepts in the context of process model comprehension, e.g., [11]. Our assessment of WM will be in line with [13].

EF, seen as the cognitive control processes that regulate thought and action, are represented by multiple correlated but separable functions [42,43]. From the perspective of cognitive psychology, executive functions regulate lower level cognitive processes and therefore shape complex performance. In general, EF play a key role for complex cognitive activities [21], [42], [44], in a variety of work-related tasks, e.g.,

[21], [45,46], and EF are essential for designing process models [20]. The three most frequently investigated components are *response inhibition* (inhibition), *updating working memory representations* (updating), and *set shifting* (shifting) [42]. Inhibition therefore, is seen as the ability to inhibit dominant, automatic responses. Updating, on the other hand describes the ability to appropriately update incoming information of relevance for the task at hand by replacing old, now irrelevant information with newer, relevant information. Shifting describes the ability to flexibly switch back and forth between tasks or different mental sets. For instance, [20], -in line with [47]-, stresses the importance of EF for the process of creating process models in terms of attentional control, goal maintenance and suppression of distracting information, error monitoring, and effortful memory search. Our psychometric assessment of EF will go in line with [42] and [48].

Process model quality is used as dependent construct for our research. In line with [49], we are going to consider *syntactic errors* (e.g., violations of the soundness property) and *semantic errors* as quality measures of the process model. Semantic errors are referring to the validity of the model (i.e., statements within the model are correct and related to the domain) and completeness (i.e., all relevant and correct statements to solve a problem are contained by the model). We will utilize existing automated techniques, e.g., [50] to quantify syntactic errors. For assessing semantic quality, due to the absence of a fully automated approach [17], we will apply a semi- automated approach. In addition, expert assessments in form of an iterative consensus building process [14] will be carried out for measuring semantic quality.

3 Research Approach

In line with the theoretical background discussed above, we argue that the quality of process models strongly depends on the cognitive abilities (working memory, executive functions, reasoning ability) of the designer, as well as on the interaction of those cognitive abilities and task-specific factors resulting in CL (Fig. 1).

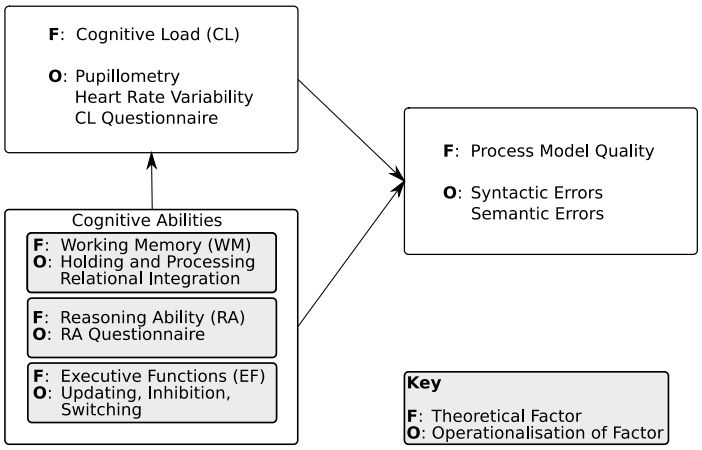

Fig. 1. Research Model

As pointed out by [20], cognitive abilities such as *reasoning ability*, *working memory*, and *executive functions* are crucial for designing process models of high quality, and for design activities, and problem solving in general [13].According to the research model we state our first research question:

Q 1: Designer's cognitive abilities, namely working memory, executive functions, and reasoning ability positively predict process model quality.

Cognitive load is described as demands on the cognitive system imposed by a task and depends on the available cognitive resources [21]. Subjects with higher working memory capacity executing the same task show lower cognitive load [21]. Reasoning ability, in turn, helps to link strategies to goals freeing resources of working memory [27, 28, 29], thus, leading to reduced cognitive load. Further, executive functions are seen as cognitive control processes regulating thought and action [21], which ensure task focus, and therefore, reduce cognitive load. According to our research model we state our second and third research question as follows:

Q 2: Higher cognitive load during task execution leads to lower process model quality.
Q 3: Within subjects executing the same modeling task, higher individual capabilities regarding working memory, executive functions and reasoning ability lead to lower cognitive load.

While subjects execute modeling tasks containing a variety of different task-specific factors cognitive load is imposed on the designer's cognitive system, e.g., [21]. Depending on the level of demands imposed by a specific part of the modeling task (e.g., building of a mental model) different amounts of cognitive load should be apparent. This leads to our fourth research question:

Q 4: Which task-specific factors within the process of designing process models are most demanding?

Currently, we are finishing the planning-phase of our research and within the next months we will carry out data collection. We are going to assess the constructs described above on 80 subjects without previous modeling experience to control the influence of former process modeling knowledge. This research is going to shed some light on the cognitive requirements for creating process models and how cognitive abilities, namely working memory, executive functions, and reasoning ability and the resulting cognitive load imposed by the interaction of those cognitive abilities and the modeling process itself affect the creation of process models and the achieved process model quality. Moreover, by applying continuous measurements of cognitive load, the identification of the most demanding factors within the process of creating process models should be possible, giving advice to the improvement of modeling notations and associated tool support.

References

1. Burton-Jones, A., Meso, P.: The Effects of Decomposition Quality and Multiple Forms of Information on Novices' Understanding of a Domain from a Conceptual Model. Journal of the Association for Information Systems 9(12), Article 1 (2008)
2. Recker, J., Rosemann, M., Indulska, M., Green, P.: Business Process Modeling- A Comparative Analysis. Journal of the Association for Information Systems 10(4), Article 1 (2009)
3. Fettke, P.: How Conceptual Modeling Is Used. Communications of the Association for Information Systems 25(1), Article 43 (2009)
4. Mendling, J.: Metrics for Process Models (Vol. 6). Springer, Berlin, Heidelberg (2008)
5. Weber, B., Reichert, M., Mendling, J., Reijers, H.A.: Refactoring Large Process Model Repositories. Computers in Industry 62(5), 467–486 (2011)
6. Reijers, H. A., Mendling, J.: A Study Into the Factors That Influence the Understandability of Business Process Models. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 41(3), 449–462 (2011)
7. Mendling, J., Strembeck, M., Recker, J.: Factors of Process Model Comprehension - Findings from a Series of Experiments. Decision Support Systems 53(1), 195–206 (2012)
8. Mendling, J., Reijers, H. A., Recker, J.: Activity Labeling in Process Modeling: Empirical Insights and Recommendations. Information Systems 35(4), 467–482 (2010)
9. Figl, K., Recker, J., Mendling, J.: A study on the Effects of Routing Symbol Design on Process Model Comprehension. Decision Support Systems 54(2), 1104–1118 (2013)
10. Figl, K., Mendling, J., Strembeck, M.: The Influence of Notational Deficiencies on Process Model Comprehension. Journal of the Association for Information Systems 14(6), Article 1 (2013)
11. Recker, J., Reijers, H.A., Van de Wouw, S.G.: Process Model Comprehension: The Effects of Cognitive Abilities, Learning Style, and Strategy. Communications of the Association for Information Systems 34(9), 199–222 (2014)
12. Pinggera, J., Zugal, S., Weber, B., Fahland, D., Weidlich, M., Mendling, J., Reijers, H.A.: How the Structuring of Domain Knowledge Helps Casual Process Modelers. In: Parsons, J., Saeki, M., Shoval, P., Woo, C., Wand, Y., (eds.) Conceptual Modeling – ER 2010, pp. 445–451. Springer, Berlin, Heidelberg (2010)
13. Sachse, P., Ulich, E.: Psychologie Menschlichen Handelns: Wissen & Denken - Wollen & Tun. Pabst Science Publishers, Lengerich (2014)
14. Recker, J., Safrudin, N., Rosemann, M.: How Novices Design Business Processes. Information Systems 37(6), 557–573 (2012)
15. Figl, K., Recker, J.: Exploring Cognitive Style and Task-specific Preferences for Process Representations. Requirements Engineering (2014)
16. Simon, H. A.: The Sciences of the Artificial – 3rd Edition. The MIT Press, Cambridge, Mass (1996)
17. Soffer, P., Kaner, M., Wand, Y.: Towards Understanding the Process of Process Modeling: Theoretical and Empirical Considerations. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) Business Process Management Workshops, pp. 357–369. Springer, Berlin, Heidelberg (2012)
18. Cinaz, B.: Monitoring of Cognitive Load and Cognitive Performance Using Wearable Sensing. Dissertation No. 21091, ETH Zurich, Zurich, Switzerland (2013)
19. Wickens, C.D., Hollands, J.G., Parasuraman, R., Banbury, S.: Engineering Psychology & Human Performance – 4th edition. Pearson, Boston (2012)

20. Wilmont, I., Hengeveld, S., Barendsen, E., Hoppenbrouwers, S.: Cognitive Mechanisms of Conceptual Modelling. In: Ng, W., Storey, V.C., Trujillo, J.C. (eds.) Conceptual Modeling, pp. 74–87. Springer, Berlin, Heidelberg (2013)
21. Boduroglu, A., Minear, M., Shah, P.: Working Memory. In: Durso, F.T., Nickerson, R.S., Dumais, S.T., Lewandowsky, S., Perfect, T.J. (eds.) Handbook of Applied Cognition – 2nd Edition, pp. 55–82. John Wiley & Sons, Chichester, England (2007)
22. Lin, T., Imamiya, A., Mao, X.: Using Multiple Data Sources to Get Closer Insights into User Cost and Task Performance. Interacting with Computers 20(3), 364–374 (2008)
23. Novak, D., Beyeler, B., Omlin, X., Riener, R.: Workload Estimation in Physical Human-Robot Interaction Using Physiological Measurements. Interacting with Computers (2014).
24. Sternberg, R.J., Kaufman, S.B. (eds.) The Cambridge Handbook of Intelligence. Cambridge University Press, Cambridge (2011)
25. Zugal, S., Pinggera, J., Reijers, H., Reichert, M., Weber, B.: Making the Case for Measuring Mental Effort. In: Proceedings of the Second Edition of the International Workshop on Experiences and Empirical Studies in Software Modelling, pp. 6:1–6:6. ACM, New York, NY, USA (2012)
26. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. North-Holland Press, Amsterdam (1988)
27. Gobet, F., Waters, A.J.: The Role of Constraints in Expert Memory. Journal of Experimental Psychology - Learning, Memory, and Cognition 29(6), 1082–1094 (2003)
28. Horn, J., Masunaga, H.: A Merging Theory of Expertise and Intelligence. In: The Cambridge Handbook of Expertise and Expert Performance (pp. 587–611). Cambridge University Press, New York, NY (2006)
29. Proctor, R.W., Vu, K.-P.L.: Laboratory Studies of Training, Skill Acquisition, and Retention of Performance. In: Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R.R. (eds.) The Cambridge Handbook of Expertise and Expert Performance, pp. 265–286. Cambridge University Press, Cambridge, UK (2006)
30. Raven, J.J.: Raven Progressive Matrices. In: McCallum, R.S. (ed.) Handbook of Nonverbal Assessment, pp. 223–237. Springer US (2003)
31. Burgess, G.C., Gray, J.R., Conway, A.R.A., Braver, T.S.: Neural Mechanisms of Interference Control Underlie the Relationship between Fluid Intelligence and Working Memory span. Journal of Experimental Psychology - General 140(4), 674–692 (2011)
32. Cowan, N., Elliott, E.M., Scott Saults, J., Morey, C.C., Mattox, S., Hismjatullina, A., Conway, A.R.A.: On the Capacity of Attention: Its Estimation and its Role in Working Memory and Cognitive Aptitudes. Cognitive Psychology 51(1), 42–100 (2005)
33. Just, M.A., Carpenter, P.A.: A Capacity Theory of Comprehension: Individual Differences in Working Memory. Psychological Review 99(1), 122–149 (1992)
34. Kyllonen, P.C., Stephens, D.L.: Cognitive Abilities as Determinants of Success in Acquiring Logic Skill. Learning and Individual Differences 2(2), 129–160 (1990)
35. Hambrick, D.Z., Engle, R.W.: Effects of Domain Knowledge, Working Memory Capacity, and Age on Cognitive Performance: An Investigation of the Knowledge-Is-Power Hypothesis. Cognitive Psychology 44(4), 339–387 (2002)
36. Oberauer, K., Süss, H.-M., Schulze, R., Wilhelm, O., Wittmann, W.W. Working Memory Capacity - Facets of a Cognitive Ability Construct. Personality and Individual Differences 29(6), 1017–1045 (2000)
37. Oberauer, K., Süβ, H.-M., Wilhelm, O., Wittmann, W.W.: Which Working Memory Functions Predict Intelligence? Intelligence 36(6), 641–652 (2008)

38. Daneman, M., Carpenter, P.A.: Individual Differences in Working Memory and Reading. Journal of Verbal Learning and Verbal Behavior 19(4), 450–466 (1980)
39. Pinggera, J., Zugal, S., Weber, B., Fahland, D., Weidlich, M., Mendling, J., Reijers, H.A.: How the Structuring of Domain Knowledge Helps Casual Process Modelers. In: Parsons, J., Saeki, M., Shoval, P., Woo, C., Wand, Y., (eds.) Conceptual Modeling – ER 2010, pp. 445–451. Springer, Berlin, Heidelberg (2010)
40. Zugal, S., Pinggera, J., Weber, B.: Assessing Process Models with Cognitive Psychology. In: EMISA 2011, pp. 177-182 (2011)
41. Claes, J., Gailly, F., Poels, G.: Cognitive Aspects of Structured Process Modeling. In: Franch, X., Soffer, P. (eds.) Advanced Information Systems Engineering Workshops, pp. 168–173. Springer, Berlin, Heidelberg (2013)
42. Friedman, N.P., Miyake, A., Young, S.E., Defries, J.C., Corley, R.P., Hewitt, J.K.: Individual Differences in Executive Functions are Almost Entirely Genetic in Origin. Journal of Experimental Psychology - General 137(2), 201–225 (2008)
43. Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D.: The Unity and Diversity of Executive Functions and their Contributions to Complex "Frontal Lobe" Tasks: a Latent Variable Analysis. Cognitive Psychology 41(1), 49–100 (2000)
44. Kane, M.J., Engle, R.W.: The Role of Prefrontal Cortex in Working-memory Capacity, Executive Attention, and General Fluid Intelligence: an Individual-differences Perspective. Psychonomic Bulletin & Review 9(4), 637–671 (2002).
45. Ackerman, P.L.: A Correlational Analysis of Skill Specificity: Learning, Abilities, and Individual Differences. Journal of Experimental Psychology: Learning, Memory, and Cognition 16(5), 883–901 (1990)
46. Reder, L.M., Schunn, C.D.: Bringing Together the Psychometric and Strategy Worlds: Predicting Adaptivity in a Dynamic Task. In: Gopher, D., Koriat´, A. (eds.) Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of Theory and Application, pp. 315–342. The MIT Press, Cambridge, MA, US (1999)
47. Engle, R.W., Kane, M.J., Tuholski, S.W.: Individual Differences in Working Memory Capacity and What they Tell us about Controlled Attention, General Fluid Intelligence, and Functions of the Prefrontal Cortex. In: Miyake, A., Shah, P. (eds.) Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, pp. 102–134. Cambridge University Press (1999)
48. Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., Klauer, K.C.:. Behavioral Components of Impulsivity. Journal of Experimental Psychology - General 143(2), 850–886 (2014)
49. Krogstie, J., Sindre, G., Jørgensen, H.:. Process Models Representing Knowledge for Action: a Revised Quality Framework. European Journal of Information Systems 15(1), 91–102 (2006)
50. Van der Aalst, W.M.P.: Verification of Workflow Nets. In: Azéma, P., Balbo, G. (eds.) Application and Theory of Petri Nets, pp. 407–426. Springer, Berlin, Heidelberg (1997).

An Evolutionary Explanation of Graph Comprehension using fMRI

Roozmehr Safi, Eric Walden, Gabriel Cogo, David Lucus, and Elshan Moradiabadi

Texas Tech Rawls, College of Business, Lubbock, Texas
{r.safi, eric.walden, gabriel.cogo, david.lucus, elshan.moradiabadi}@ttu.edu

Abstract. Evolution has equipped Homo sapiens with a wide range of inherent abilities. One of those abilities is comprehending graphical representations. We claim that comprehension is only inherent if the representation has an analogy in the evolutionary environment. We test this using a fMRI study to show that certain graphs activate the visual cortex and others do not. Furthermore those that activate the visual cortex result in greater accuracy.

Keywords: fMRI · graphicacy · evolution.

1 Introduction

Graphs are an important way to represent data. The idea behind graphs is that they represent numerical data in a visual manner so that people can take advantage of the 20 percent of the cortex devoted to visual processing [1]. However, to properly take advantage of the oversized visual cortex we must construct graphs the right way. Specifically, we must construct graphs so that the cortex can process them natively. By natively, we mean in they need to represent data in the way the environment represents visual stimuli.

In this paper we compare one way in which graphs can be constructed natively—3-D surface graphs—to one non-native graph construction—balloon race graphs. We propose that native 3-D graphs will encourage more processing in the visual cortex and hence result in more accurate reading of the graphs.

2 Theory

For all organisms the genetic code specifies a blueprint for building the organism. This includes a blueprint for the structure of the brain. The blueprint for all Homo sapiens brains is basically the same. Some may have slight variation, and some may even have large defects, but if the genetic blueprint is too different, then we are not looking at a homo sapiens any more. Chimpanzees have about 96% of the same in-

structions in their genetic blueprint as Homo sapiens do, so changing 4% of the instructions builds a completely different animal.

This genetic blueprint is important because the brain is built not as an undifferentiated mass of wrinkly stuff, but as an integrated set of specific problem solving structures. For example, the part of you brain that controls body movements runs roughly from your ears to the top of your head. It contains the largest neurons in the central nervous system, called Betz cells. These neurons are not only large, but bushy reaching in every direction. This makes sense for controlling the various muscles of the body because each muscle needs to work with each other muscle, so the controllers need to communicate with each other. This interlinking allows us to make seamless motion. On the other hand in the primary visual cortex neurons are smaller and more tightly packed, with long tails. That is because seamless vision is bad, it is really the seems that define vision. Moreover, there are many more photons hitting the eye than there are muscles to control, so more small neurons are better for processing visual information. These areas are built in specific ways with specific structures by or genetic blueprint to solve specific problem.

What sort of problem solving structures can be encoded into the genetic blueprint? To understand that we have to understand how evolution shapes our genetic blueprint. In each generation, the genetic blueprint builds a set of people. Those people have some genetic differences, which translate into differences in the type of person they are. Some are taller, some are darker, and some have a larger putamen because of a single change of one DNA base pair called [2]. As these people grow and live they face challenges, which they must overcome to thrive. Those that thrive, produce offspring which have similar genetic make up, and those that do not thrive produce fewer offspring, or no offspring if failure to thrive results in death.

These sorts of problems are called adaptive problems [3], and if many generations of any organism face them then the genetic code of the species as a whole shifts. For example, Tawny owls in Finland are either brown or white. Over time as arctic ice in Finland has melted and white snow has been replaced by brown dirt and trees, the species as a whole has become more brown. Similarly, in the 1800's peppered moths as a species were predominately a specked black and white pattern that matched the bark on pristine trees in England. However, with the onset of the industrial revolution, trees around industrial cities such as Manchester became black with soot, and by the 1900's most peppered moths were solid black.

Like the coloration of moths and owls, the structures of the human brain were formed as solutions to adaptive problems. Thus, we expect that a good data visualization will be one that simulates some sort of problem faced in the natural environment, and a poor data visualization will be on that does not take advantage of any of the specially designed structures in the brain, because it does not resemble anything our ancestors might have been expected to see and have to deal with.

One example of something that might be a great data visualization is the Wong-Baker face pain rating scale. Certainly, our ancestors faced the problem of recognizing emotions on faces. The Wong-Baker scale taps into brain structures specifically designed to interpret facial expressions. You can take Wong-Baker to anyone in the

world, from Stone Age tribes in the Amazon to a French Baker to a Chinese motorcycle mechanic and they intrinsically understand what it means.

Evolutionary theory suggests that graphs based on visual patterns that were present in the ancestral environment will be more effectively evaluated than graphs that do not correspond to anything one may have expected to encounter in the ancestral environment. Moreover, the better performance will be a direct result of using brain regions in the visual cortex specifically designed to understand visual stimuli. This area is called the ventral stream and it runs from the posterior part of the brain, in an inferior direction. In laymen's terms, from the back of the head down to the bottom of the brain.

3 Methods

To test our theory we used two sorts of graphs: one that matched the types of visual problems the brain was designed to solve, and one that did not. For the type of graph that the brain is designed to comprehend we used 3-D surface graphs. The brain should obviously be able to deal with this in a visual manner because the environment is composed of 3-D surfaces. Sometimes we forget this living in modern right-angles buildings, but the natural environment in which Homo sapiens evolved was very much filled with 3-D surfaces.

For non-native processing we choose balloon race graphs. These are scatter plots where the size of the markers represent a 3rd dimension and the color of the markers represent a 4th dimension. These are a popular type of graph, but there is no natural analog to a balloon race graph. Both of these are shown in the figure below.

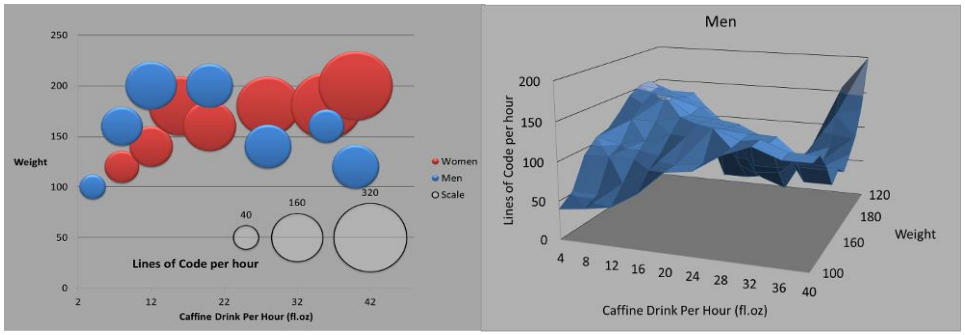

Fig. 1. Balloon race graph (right) and 3-D surface graph (left)

Subjects viewed 30 of each kind of graph while in a fMRI scanner and were asked to answer multiple-choice questions about the graphs (4 choices). Half of the questions were in sample and half were out of sample. The questions were randomized so that the same question was put to a different graph for different people.

3.1 Subjects

Subjects were 17 students at a university there were 9 males, 15 right-handers and the average age was 20.9.

3.2 Data analysis

Data were collected on a 3 Tesla machine with a repetition time of 2500 ms, echo time of 30 ms, and flip angle of 90 degrees. Data was high passed filtered at 100 s. Motion was corrected using FSL's Flirt [4,5]. Functional scans were registered to high resolution T1 scans and then to MNI 152 standard space. A linear model was created with dummy variables for fixation, 3-D graph and balloon race graph times. This was then convolved with a gamma function to account for the delay in the hemodynamic response function. All subjects were grouped using a random effects model.

4 Results

We found is that subjects answered the questions correctly 52.4% of the time when using 3-D graphs and only 35.7% of the time when using balloon race graphs. This difference was significant in a logistic regression (Chi-square stat 1 df = 19.61, p-value <0.0001). Thus, people were significantly better at answering 3-D surface graphs than balloon race graphs.

When we contrasted the brain activation when answering questions about the 3-D surface graphs to the brain activation when answering questions about balloon race graphs, we found significant bilateral activation in the ventral stream. By contrast, the balloon race graphs produced a random pattern of activation in a variety of non-visual areas.

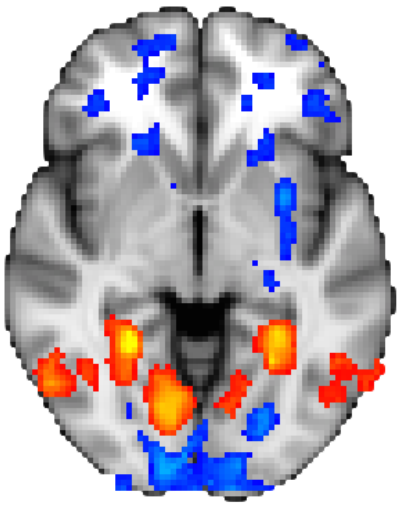

Fig. 2. Areas of greater activation for 3-D graphs in red and for Balloon graphs in blue. Slice at Y = -34 in MNI152 coordinates.

5 Discussion

When looking at graphs that are similar to things our ancestors faced in their native environment, people are much better at answering questions about those graphs. They have a lot less activation (about a quarter as much), but they activate the right structures—structures that evolution wrote into our genetic code to solve that sort of prob-

lem. However, when looking at graphs with no analog in the native environment people had a difficult time answering questions, though they recruited much more of their brain. Unfortunately, without specifically designed structures to comprehend the graphs, people were like dogs trying to turn a doorknob. They are just not built for it.

Just like a dog trying to turn a doorknob, with a lot of training an exceptional individual can learn to use graphs they are not designed to understand. However, even the most exceptional dog cannot turn a doorknob better than an average person. The whole point of visualizations is to share them with average people in an intuitive way.

Unfortunately, the proliferation of visualizations is probably misplaced. It is probably the novelty of these visualizations that make them exciting. Many visualizations are beautiful artistic works. But art is notoriously ambiguous, and if the goal is understanding, then we need to be vary cautious. A good visualization is one that takes advantage of the neural capabilities evolution endowed us with. A good data visualization is one that mimics something our ancestors faced in their environment over many generations.

References

1. Wandell, B.A., Dumoulin, S.O., Brewer, A.A.: Visual Cortex in Humans. Encyclopedia of Neuroscience 10, 251-257 (2009)
2. Hibar, D.P., Stein, J.L., Renteria, M.E., Arias-Vasquez, A., Desrivières, S., Jahanshad, N., Toro, R., Wittfield, K., Abramovic, L., Andersson, M., Aribisala, B.S., Armstrong, N.J., Bernard, M., Bohlken, M.M., Boks, M.P., Bralten, J., Brown, A.A., Chakravarty, M.M., Chen, Q., Ching, C.R.K., et al.:. Common Genetic Variants Influence Human Subcortical Brain Structures. Nature (2015)
3. Cosmides, L., Tooby, J.: Beyond Intuition and Instinct Blindness: Toward an Evolutionarily Rigorous Cognitive Science. Cognition 50(1), 41-77 (1994)
4. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. Neuroimage 17(2), 825-841 (2002)
5. Jenkinson, M., Smith, S.: A Global Optimisation Method for Robust Affine Registration of Brain Images. Medical Image Analysis 5(2), 143-156 (2001)

Investigation of the relationship between visual website complexity and users' mental workload: A NeuroIS perspective

Ricardo Buettner
FOM University of Applied Sciences, mis-Institute, Munich, Germany
ricardo.buettner@fom.de

Abstract. We report promising research-in-progress results from an ongoing experiment on the relationship between visual website complexity and users' mental workload. Applying pupillary based workload assessment as a NeuroIS methodology we found indications that navigation complexity, i.e., the number of (sub)menus, is more problematic than information complexity.

Keywords: NeuroIS · eye-tracking · mental · workload · pupillary diameter · IS complexity · website complexity · navigation complexity · information complexity

1 Introduction

Website/webpage complexity affects a user's mental workload [1]. Huang [2] identified the amount of information and the number of links as important attributes of website complexity. The problem from a website design perspective is how to balance the dilemma of a complex menu structure (a lot of menu links and submenus) but non-complex pieces of information or a non-complex menu structure (with fewer links/submenus) with a high amount of information (more complex). To evaluate this problem researchers need a convenient way to assess a user's mental workload. Determining a user's mental workload is often mentioned as a fundamental problem in IS research (e.g. [3,4]) from various theoretical perspectives (e.g. cognitive load, task technology fit, job demands-resources), particularly in NeuroIS (e.g. [5-10]).

In recent years very interesting results have emerged from a new field called NeuroIS in which efforts have been made to determine a user's mental workload based on objective psychophysiological measurements [8-10]. IS scholars have used a pupillary based mental workload assessment already using realistic experimental setups, e.g., route planning [11,12], E-mail classification [11], decision support systems [13], and social networks [6,14].

To the best of our knowledge there is no study investigating the relationship between visual website complexity and users' mental workload using psychophysiological measures – with one exception: The work of Wang et al. [1] investigated website complexity from a cognitive load perspective via eye-tracking technology. Using fixation count and fixation duration they found increased fixation counts, fixation

durations and task completion times when performing simple tasks. Interestingly they did not analyze pupillary measures in order to evaluate mental workload.

That is why we study the usage of three website variants with systematic manipulations of navigation and information complexity using eye-tracking based pupillary diameter responses. With our work we contribute to IS complexity research. In addition, we address a very practical problem for website designers.

2 Methodology

2.1 Applying the NeuroIS guidelines

In order to clearly contribute to NeuroIS research and show strong methodological rigor, we strictly followed the NeuroIS guidelines of vom Brocke et al. [15]. In particular, to assess prior research in the field of measuring mental workload as an important IS construct, a comprehensive literature review was conducted (cf. [16]). To base our experimental design adequately on solid research in related fields of neuroscience [15] we reviewed the fundamental anatomic mechanism of the pupillary dilation controlled by the vegetative nervous system and the key role of the Edinger-Westphal nucleus that is inhibited by mental workload and directly leads to a pupillary dilation. Our methodology uses eye-tracking-based pupillometry as a well-established approach in physiology and psychology “widening the 'window' of data collection" [17, p. 93]. With our method, bio-data (i.e. pupil diameter) can be used to better understand mental workload as an IS construct (cf. guideline 4 of [15]). In comparison to other neuroscience tools eye-tracking-based pupillometry is the contact-free and efficient method of choice [18]. We applied the guidelines and standards from Duchowski [19] and the Eyegaze Edge™ manual.

2.2 Measurements

To capture the pupillary diameter as the measure of interest in this research, eye-tracking was performed using the binocular double Eyegaze EdgeTM System eye-tracker paired with a 19" LCD monitor (86 dpi) set at a resolution of 1280x1024, whereby the eye-tracker samples the pupillary diameter at a rate of 60 Hz for each eye separately.

2.3 Stimuli

Following [20] we manipulated website visual complexity via the number of links in our menu structure (resp. submenus). According to [1] we chose three contrary but balanced levels for navigation and information complexity. Navigation complexity was manipulated by the (sub)menu structure (low: 3 menus; average: 3×3 (sub)menus; high: $3 \times 3 \times 3$ (sub)menus). Information complexity was manipulated by content/text partitioning. All three variants (system A,B,C; see Figure 1) contained the same content/information in summary but we divided this content into (sub)menu-

specific pieces of information. Luminescence levels of the three systems variants were checked.

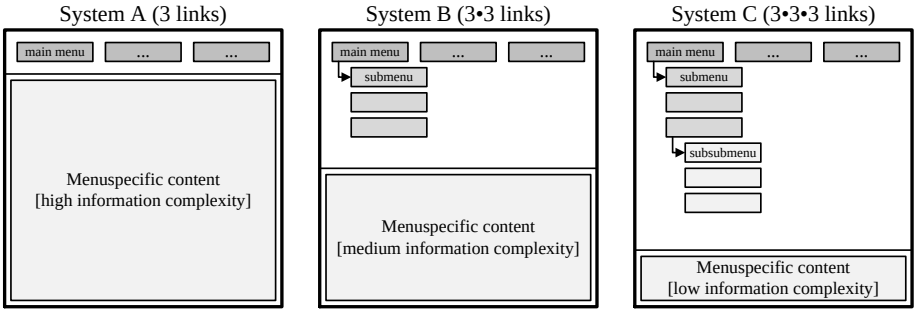

Fig. 1. Conceptualized website complexity (System A: low menu complexity – high information complexity; System B: average menu complexity – average information complexity; System C: high menu complexity – low information complexity)

Please note that we directly tested objective website complexity, since perceived website complexity correlated only medially with objective website complexity (r = .3 according to [21, p. 515], cf. [22]).
The participants in our experiment had to perform nine distinctive search tasks – three for each system. In order to counter-balance our design, the test order of the systems (A, B, C) was randomized. In addition, for every test system (A, B, C) three of the nine search tasks were randomly assigned.

2.4 Description of the test procedure

Prior to all data collection each test participant is welcomed by the experimenter (the supervisor of the experiment). After that the participant has to fill out a consent form and also a questionnaire with demographics (stage 1). In stage 2, we take the necessary precautions for the experiment during which we make use of the eye-tracker. Hence, the eye-tracker is calibrated. In stage 3, the experiment starts with the first search task the participant has to accomplish.

2.5 Data cleansing

Only naturally determined artifacts, e.g. by eye-blinks, were deleted.

3 Results

3.1 Sample Characteristics

Our 13 participants were aged from 23 to 35 years (M=28.1, S.D.=3.9). 6 persons were female, 7 male.

3.2 Relationship between visual website complexity and users' mental workload

We found clear pupil diameter differences between the three system variants (table 1) which were partly significant already at this stage of research (n = 13, A/B: $p_{left\ eye} < .05$, $p_{right\ eye} < .05$; B/C: $p_{left\ eye} < .1$, $p_{right\ eye}$ n.s.; A/C: $p_{left\ eye} < .01$, $p_{right\ eye} < .01$):

Table 1. Mean of pupillary diameters in relation to system variant

System	PD [mm]	
	left eye	right eye
System A	3.220	3.249
System B	3.246	3.278
System C	3.279	2.290

4 Discussion, limitations and future research

From a mental workload perspective the system A is the model of choice since the pupillary based mental workload indicator is lowest for this system variant.

That means for the practical website design perspective that complex menu structures with a lot of menu links and submenus should be avoided. Instead, the designers should use fewer submenus (lower navigation complexity) but more text (more information complexity) – contingently with scroll bars.

From a theoretical point of view our work contributes to IS complexity research. Our results indicate that navigation complexity (i.e., the number of (sub)menus) is more problematic than information complexity from a mental workload perspective.

At this stage of the research our main limitation is rooted in the small sample size (n=13). In the future we will test more participants (n~70). In addition, in an extended version of this paper we will report on triangulated NASA TLX evaluations and results from electrodermal activity assessments for the whole sample.

References

1. Wang, Q., Yang, S., Liu, M., Cao, Z., Ma, Q.: An Eye-tracking Study of Website Complexity from Cognitive Load Perspective. Decision Support Systems 62, 1-10 (2014)
2. Huang, M.H.: Designing Website Attributes to Induce Experiential Encounters. Computers in Human Behavior 19(4), 425-442 (2003)
3. Stassen, H.G., Johannsen, G., Moray, N.: Internal Representation, Internal Model, Human Performance Model and Mental Workload. Automatica 26(4), 811-820 (1990)
4. Johannsen, G., Levis, A.H., Stassen, H.G.: Theoretical Problems in Man-machine Systems and Their Experimental Validation. Automatica 30(2), 217-231 (1992)
5. Buettner, R.: Analyzing Mental Workload States on the Basis of the Pupillary Hippus. In: Proceedings of NeuroIS 2014, p. 52 (2014)

6. Buettner, R., Sauer, S., Maier, C., Eckhardt, A.: Towards ex ante Prediction of User Performance: A novel NeuroIS Methodology based on Real-Time Measurement of Mental Effort. In: Proceedings of HICSS-48, pp. 533-542. AIS (2015)
7. De Guinea, A.O., Titah, R., Léger, P.M.: Explicit and Implicit Antecedents of Users' Behavioral Beliefs in Information Systems: A Neuropsychological Investigation. Journal of Management Information Systmes 30(4), 179-210 (2014)
8. Dimoka, A., Pavlou, P.A., Davis, F.D.: NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. Information Systems Research 22(4), 687-702 (2011)
9. Dimoka, A., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P.H., Pavlou, P.A., Müller-Putz, G., Riedl, R., vom Brocke, J., Weber, B.: On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. MIS Quarterly 36(3), 679-A19 (2012)
10. Riedl, R., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Dimoka, A., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Müller-Putz, G., Pavlou, P.A., Straub, D.W., vom Brocke, J., Weber, B.: On the Foundations of NeuroIS: Reflections on the Gmunden Retreat 2009. Communications of the Association for Information Systems 27, 243-264 (2010)
11. Iqbal, S.T., Adamczyk, P.D., Zheng, X.S., Bailey, B.P.: Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution. In: Proceedings of CHI 2005, pp. 311-320 (2005)
12. Bailey, B.P., Iqbal, S.T.: Understanding Changes in Mental Workload during Execution of Goal-Directed Tasks and Its Application for Interruption Management. ACM Transactions on Computer-Human Interaction 14(4), Article 21 (2008)
13. Buettner, R.: Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology. In: KI 2013: Advances in Artificial Intelligence - LNCS 8077, pp. 37-48. Springer, Berlin, Heidelberg (2013)
14. Buettner, R., Daxenberger, B., Eckhardt, A., Maier, C.: Cognitive Workload Induced by Information Systems: Introducing an Objective Way of Measuring based on Pupillary Diameter Responses. In: Proceedings of Pre-ICIS HCI/MIS, Paper 20 (2013)
15. Vom Brocke, J., Liang, T.P.: Guidelines for Neuroscience Studies in Information Systems Research. Journal of Management Information Systems 30(4), 211-234 (2014)
16. Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: Proceedings of ECIS 2009, pp. 2206-2217. AIS (2009)
17. Goldinger, S.D., Papesh, M.H.: Pupil Dilation Reflects the Creation and Retrieval of Memories. Current Directions in Psychological Science 21(2), 90-95 (2012)
18. Laeng, B., Sirois, S., Gredebäck, G.: Pupillometry: A Window to the Preconscious? Perspectives on Psychological Science 7(1), 18-27 (2012)
19. Duchowski, A.T.: Eye Tracking Methodology: Theory and Practice (2nd Edition). Springer, London, UK (2007)
20. Deng, L., Poole, M.S.: Affect in Web Interfaces: A Study of the Impacts of Web Page Visual Complexity and Order. MIS Quarterly 34(4), 711-A10 (2010)
21. Nadkarni, S., Gupta, R.: A Task-Based Model of Perceived Website Complexity. MIS Quarterly 31(3), 501-524 (2007)
22. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Lawrence Erlbaum, Hillsdale, NJ, USA (1988)

Measuring Cognitive Load During Process Model Creation*

Barbara Weber, Manuel Neurauter, Jakob Pinggera, Stefan Zugal, Marco Furtner, Markus Martini, and Pierre Sachse

University of Innsbruck, Austria
{barbara.weber, manuel.neurauter, jakob.pinggera, stefan zugal, marco.furtner, markus.martini, pierre.sachse}@uibk.ac.at

Abstract. While factors impacting process model comprehension are relatively well understood by now, little is known about process model creation and factors impacting the quality of the resulting process model as well as the modeler's cognitive load. In this paper we propose to combine a continuous, psycho–physiological measurement of cognitive load with a detailed analysis of the modeler's interactions of the modeling environment as well as eye movement analysis to obtain task–specific imposed cognitive load values. We present initial results in terms of a tool, lessons learnt from a pilot study and discuss upcoming challenges. This work provides the basis for investigating task imposed cognitive load during process model creation by enabling a dynamic, semi–automatic analysis of cognitive load.

Keywords: business process management · process modeling · process model creation · eye tracking · cognitive load

1 Introduction

Nowadays, business process modeling is heavily used in various business contexts. For instance, process models help to obtain a common understanding of a company's business processes [1], facilitate inter–organizational business processes [2], and support the development of information systems [3]. Still, process models in industrial process model collections often display a wide range of quality problems [4], calling for a deeper investigation of process model quality.

Previous research activities resulted in a good understanding on factors impacting process model comprehension. For instance, notational deficiencies [5], modeling expertise [6], and process knowledge [7] have shown to provide

* Modeling Mind is a collaborative research effort of the Institute of Computer Science and the Institute of Psychology at the University of Innsbruck. Modeling Mind is funded by Austrian Science Fund (FWF): P26609–N15.

measurable impact on the understandability of a process model. Additionally, [8] pointed out that cognitive abilities, learning style, and learning strategy provide significant impact on process model comprehension. Factors influencing process model quality in the context of process model creation, in turn, are understood to a smaller extent (e.g., [9–11]) and therefore require more attention.

Existing research on process model quality typically measures process model comprehension in terms of accuracy (the number of correct answers about models, e.g., [5,12,13]) and answering speed [5,12]. In addition, [5,13] consider cognitive load as an additional quality dimension, being measured through self– assessment. While this type of operationalization is suitable in the context of process model comprehension, it is not sufficient for studies on the creation of process models, where the cognitive demands cannot be controlled. Particularly, cognitive demands change during a modeling task considerably: For example, a model's inherent complexity (e.g., the model's size or control flow) changes during model creation whenever model elements are added or deleted.

To systematically investigate the impact of different factors on cognitive load in the context of process model creation, we propose the usage of continuous measurement of cognitive load. In particular, we aim toward a high temporal resolution by implementing psycho–physiological measurements, i.e., pupillometry. We propose to use this data for calculating task–specific cognitive load values (e.g., cognitive load for activity creation versus gateway creation). We propose a solution that uses the user interactions with the modeling environment for mapping concrete measurements in a semi–automated way to specific factors. The dynamic calculation of task–specific load values will enable data analyses that otherwise were unfeasible due to lack of experimental control. In this paper we sketch the approach, present an initial version of the developed tool, and describe initial insights of a pilot study including lessons learnt and upcoming challenges. This way, we hope to gain valuable feedback and inspiration from the research community for our next steps.

The paper is structured as follows. Section 2 illustrates our approach. Section 3 presents initial results including the tool, lessons learnt, and challenges. Section 4 concludes the paper.

2 Continuously Measuring Task–Imposed Cognitive Load

This section sketches our approach toward task–specific measurement of cognitive load: Section 2.1 elaborates on continuously measuring cognitive load, whereas Section 2.2 details on calculating task–specific cognitive load.

2.1 Continuous Cognitive Load Measurement

In general, mental effort, cognitive load, mental load, and mental workload are

often used as aliases, basically describing the same concept [14]. Cognitive load characterizes the demands of tasks imposed on the limited information processing capacity of the brain and constitutes an *individual measure* considering the individual amount of available resources [15]. While cognitive load for model comprehension tasks can be assessed easily using questionnaires [16], investigating task–imposed cognitive demands during process model creation requires more fine–grained measurements. For this, we consider continuous, psychophysiologically measurements of cognitive load, such as ocular–motoric data, pupil diameter, blink rate or heart rate variability [17]. In this work, we focus on the usage of pupil diameter as provided by table–mounted eye trackers for investigating cognitive load (an increase of the pupils' diameter is generally associated with a higher cognitive load). To enable the calculation of task–specific load values, we suggest to integrate the measurement of cognitive load, user interactions, and eye movement parameters as detailed in the next section.

2.2 Dynamic Calculation of Task Specific Cognitive Load Values

To calculate task–specific cognitive load, cognitive load measurements must be associated to a task–specific factor of interest (e.g., cognitive load associated with the creation of different types of model element). We suggest a semi–automatic approach for establishing these associations and calculating task–specific cognitive load (e.g., the average cognitive load for creating activities versus creating gateways). In particular, we suggest the usage of model interactions and eye fixations as vehicle for determining which parts of a modeling process are related to a particular aspect of process modeling. To be more specific, we assume the presence of a log of model interactions (also denoted as PPM instance) that consists of a list of events (i.e., user interactions like add activity A, add edge between start event and activity A, add gateway XOR1) with associated timeframes. Process modeling environments like Cheetah Experimental Platform (CEP) provide for such logs [18]. The log of user interactions can then, for example, be used to determine in which timeframes the modeler was working on activity creation versus gateway creation. In addition, we assume the presence of a log of fixations, comprising for each fixation additional information like timestamp and screen position. Eye fixations could be used, for example, to determine during which timeframes a user was focusing his attention on activities versus gateways. By combining model interactions and eye movement data in a single platform like CEP, we can reconstruct the model for any point in time and connect model elements with the area on the screen the subject was focusing his attention on at this particular point in time. More importantly, we obtain the data not only as part of video recordings (as in some existing software packages for eye movement analysis), but as structured data suitable for a *semi–automated* analysis.

The log of user interactions and the log of eye fixations can be filtered based on event types that are in the analysis' focus, e.g., events of type add activity or

add gateway. To assess cognitive load, timeframes are required, e.g., to calculate the average cognitive load involved in activity creation versus gateway creation. This might be done by using a sliding window with a predefined duration, which can be placed on any point in time within the PPM instance. Figure 1 illustrates how the calculation of one specific sliding window might look like.

3 Initial Results

For testing the reliability of continuous cognitive load measurements, we conducted a pilot study with three participants, i.e., two PhD students and one master student working in business process management. Each participant created a process model consisting of 19 activities, containing the basic control flow patterns: sequence, parallel split, synchronization, exclusive choice, simple merge, and structured loop [19]. As a modeling environment CEP was used, recording all model interactions. A Tobii TX300 eye tracker with 300 Hz sampling rate was used to measure pupil dilation as well as fixations.

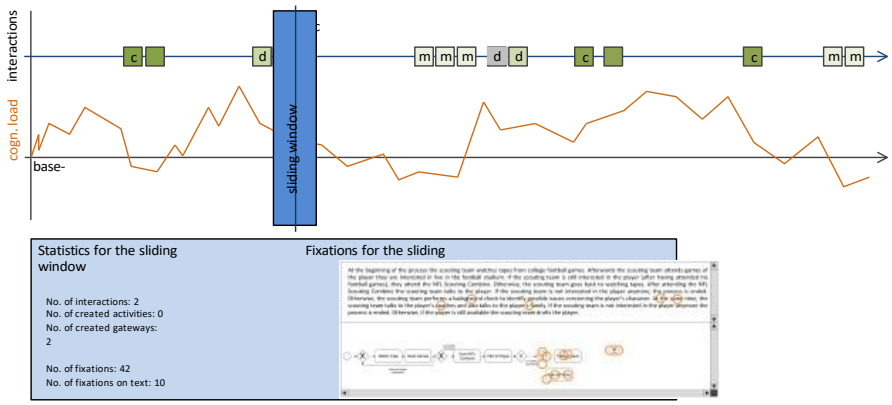

Fig. 1. Integrated PPM View

As a first step toward the calculation of task–specific cognitive load, we implemented a web application that juxtaposes cognitive load, exported from the eye tracker, with the video recording of the eye tracker. This video recording also shows the modeler's eye fixations (cf. Fig. 2).[1] Further, the user interface allows to search for phases of increased cognitive load. For this, the minimum duration of the respective phase can be set, i.e., only phases with an increased cognitive load longer than the threshold are listed.

[1] Available from: http://bpm.q-e.at/continuousMeasurement

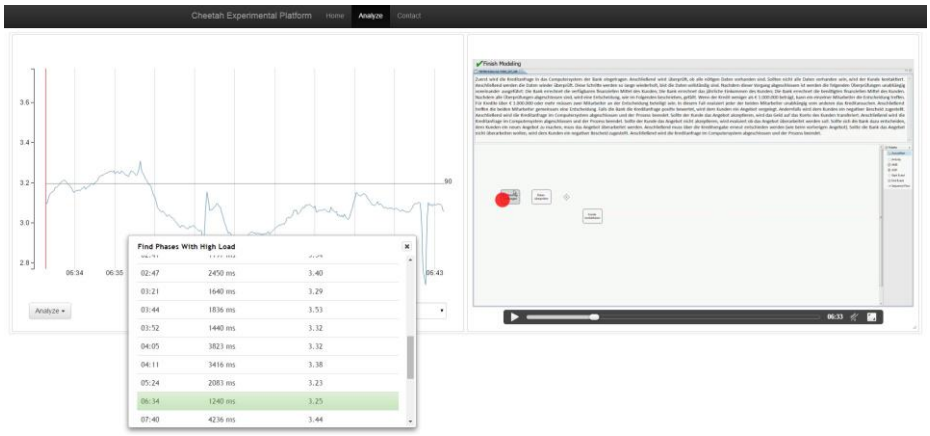

Fig. 2. CEP with Cognitive Load Analysis

We used the web application to explore the data focusing on timeframes with increased cognitive load. For one modeler we observed phases of increased cognitive load whenever this modeler had to name activities of the process model. Similar observations were not made when creating other types of nodes, e.g., XOR gateways. It seems that extracting information from the text (indicated by fixations on the textual description) and abstracting from the text to name the activity was challenging to this specific modeler. For a different modeler, we observed phases of increased cognitive load when correcting previously created parts of the model. For example, this modeler had to include a jump to a pre- vious part of the model, forcing the modeler to move some elements. This was accompanied by increased cognitive load. Further, toward the end, the modeler seemed to validate the process model. During this, the modeler changed some parts of the created model, which was accompanied by increased cognitive load. Even though we feel reinforced in pursuing this direction by the initial results, several aspects need to be considered. With respect to these, we hope for useful comments of the research community via this publication. Most importantly, we need to perform a systematic data cleaning. For instance, similar to [20], we intend to remove data fragments caused by blinks, e.g., by removing outliers larger than three times the standard deviation. Further, the creation of a process model involves motoric actions, e.g., mouse movements and typing, which might cause pupil dilation [21]. This should be considered when performing the data cleaning. Still, we are confident to obtain useful cognitive load measurements for specific timeframes of a PPM instances, since [22] successfully applied the analysis of cognitive load in a Driving Simulator—a task requiring at least the same amount of motoric actions as process modeling. Similarly, when typing, subjects might look at the keyboard, e.g., to find the appropriate finger position. Looking away from the screen and back might cause pupil reactions due to changed light conditions (dark keyboard; bright screen). Therefore, we consider complementing the analysis of pupil dilation with heart rate variability (HRV)

analysis [17] to accommodate for potentials shortcomings.
Another challenge we faced during the pilot related to baseline measurement, which we performed for conducting inter–subject comparisons. The naive assumption to ask subjects to "do nothing" incurred increased cognitive load. Therefore, we intend to utilize a dynamic baseline calculation, either immediately prior to the timeframe of interest (cf. [23]), or by averaging cognitive load for the entire duration of the modeling task (cf. [22]).

4 Summary and Outlook

This paper proposes an approach for calculating task–specific cognitive load by integrating continuous cognitive load measurements with user interactions and eye fixations. This way, paving the way for cognitive load measurement in the context of process model creation. More detailed insights into aspects of process modeling contributing to a high cognitive load might be used for giving advice to developers of new modeling notations and tool. The proposed research not only bears significant potential for process modeling research, but might be extended toward conceptual modeling as well as the design of user interfaces in general.

As for future work, we plan to work on the remaining challenges for the continuous measurement of cognitive load before addressing the calculation of task–specific cognitive load. With respect to the raised challenges, we hope to obtain valuable feedback and inspiration from the research community via this publication.

References

1. Bandara, W., Gable, G.G., Rosemann, M.: Factors and Measures of Business Process Modelling: Model Building through a Multiple Case Study. European Journal of Information Systems 14, 347–360 (2005)
2. Zhao, X., Liu, C., Yang, Y., Sadiq, W.: Aligning Collaborative Business Processes - An Organization–Oriented Perspective. Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 39, 1152–1164 (2009)
3. Dumas, M., Van der Aalst, W.M.P., ter Hofstede, A.H.M.: Process Aware In- formation Systems: Bridging People and Software Through Process Technology. Wiley–Interscience (2005)
4. Mendling, J.: Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness. Springer (2008)
5. Figl, K., Mendling, J., Strembeck, M.: The Influence of Notational Deficiencies on Process Model Comprehension. Journal of the Association for Information Systems 14,312–338 (2013)
6. Reijers, H.A., Mendling, J.: A Study into the Factors that Influence the Understandability of Business Process Models. IEEE Transactions on Systems Man and Cybernetics, Part A 41, 449–462 (2011)
7. Mendling, J., Strembeck, M., Recker, J.: Factors of Process Model Comprehen-

sion - Findings from a Series of Experiments. Decision Support Systems 53, 195–206 (2012)
8. Recker, J., Reijers, H.A., Van de Wouw, S.G.: Process Model Comprehension: The Effects of Cognitive Abilities, Learning Style, and Strategy. Communications of the Association for Information Systems 34, 199–222 (2014)
9. Mendling, J.: Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness. Springer (2008)
10. Pinggera, J., Zugal, S., Weber, B., Fahland, D., Weidlich, M., Mendling, J., Rei- jers, H.A.: How the Structuring of Domain Knowledge Can Help Casual Process Modelers. In: Proceedings of ER 2010, pp. 445-451. Springer (2010)
11. Sachse, P., Martini, M., Pinggera, J., Weber, B., Reiter, K., Furtner, M.: Das Arbeitsgedächtnis als Nadelöhr des Denkens. In: Sachse, P., Ulich, E., (eds.) Psychologie Menschlichen Handelns: Wissen & Denken - Wollen & Tun. Pabst Science Publishers (2014)
12. Figl, K., Recker, J., Mendling, J.: A Study on the Effects of Routing Symbol Design on Process Model Comprehension. Decision Support Systems 54, 1104–1118 (2013)
13. Figl, K., Laue, R.: Cognitive Complexity in Business Process Modeling. In: Proceedings of CAiSE 2012, pp. 452-466. Springer (2012)
14. Cinaz, B.: Monitoring of Cognitive Load and Cognitive Performance using Wearable Sensing. PhD thesis, ETH Zurich (2013)
15. Wickens, C.D., Hollands, J.G., Parasuraman, R., Banbury, S.: Engineering Psychology & Human Performance. Pearson Education (2012)
16. Zugal, S., Pinggera, J., Reijers, H., Reichert, M., Weber, B.: Making the Case for Measuring Mental Effort. In: Proceedings of EESSMod 2012, pp. 37-42. ACM (2012)
17. Haapalainen, E., Kim, S., Forlizzi, J.F., Dey, A.K.: Psycho–physiological Measures for Assessing Cognitive Load. In: Proceedings of Ubicomp 2010, pp. 301-310. ACM (2010)
18. Pinggera, J., Zugal, S., Weber, B.: Investigating the Process of Process Modeling with Cheetah Experimental Platform. In: Proceedings of ER–POIS 2010, pp. 13-18 (2010)
19. Van der Aalst, W.M.P., ter Hofstede, A.H.M., Kiepuszewski, B., Barros, A.P.: Workflow Patterns. Distributed and Parallel Databases 14(3), 5–51 (2003)
20. Pedrotti, M., Lei, S., Dzaack, J., Rötting, M.: A Data–driven Algorithm for Offline Pupil Signal Preprocessing and Eyeblink Detection in Low–speed Eye–tracking Protocols. Behavior Research Methods 43, 372–383 (2011)
21. Richer, F., Beatty, J.: Pupillary Dilations in Movement Preparation and Execution. Psychophysiology 22, 204–207 (1985)
22. Palinko, O., Kun, A.L., Shyrokov, A., Heeman, P.: Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. In: Proceedings of the Symposium on Eye–Tracking Research & Applications 2010, pp. 141-144. ACM (2010)
23. Klingner, J.: Fixation–aligned Pupillary Response Averaging. In: Proceedings of the Symposium on Eye–Tracking Research & Applications 2010, pp. 275-282. ACM (2010)

Cognitive Differences and their Impact on Information Perception: An Empirical Study Combining Survey and Eye Tracking Data

Lisa Falschlunger[1] , Horst Treiblmaier[1], Othmar Lehner[2], and Grabmann Elisabeth[1]

[1] University of Applied Sciences Upper Austria
{lisa.falschlunger,horst.treiblmaier, elisabeth.grabmann}@fh-steyr.at
[2] Said Business School, University of Oxford
othmar.lehner@sbs.ox.ac.uk

Abstract. Research shows that the quality of managerial decision making is dependent on both the availability and the interpretation of information. Visualizations are widely used to transform raw data into a more understandable format and to compress the constantly growing amount of information being produced. However, research in this area is highly fragmented and results are contradicting. A possible explanation for inconsistent results is the neglect of individual characteristics such as experience, working memory capacity, or cultural background. We propose a preliminary model based on an extensive literature review on cognition theory that sheds light on potential individual antecedents of information processing efficiency. Our preliminary results based on eye tracking, automated span tasks, as well as survey data show that domain expertise, spatial ability and long term orientation exert a significant influence on this cognitive construct.

Keywords: information visualization · information perception · cognitive fit · decision making · information processing efficiency

1 Introduction

The constantly growing amount of corporate data raises the question of how decision makers can effectively organize and interpret it [1]. Anticipated benefits of high volume of information are often outbalanced by the occurrence of information overload [2,3,4]. Visualizations are a means to encounter this effect. They support systematic rather than heuristic information processing [5], which is crucial to ensure quality in decision making.

Visual representation of data accelerates and improves cognition and interpretation [6], and thus improves a rational managerial decision making process [7]. According to Conati and Maclaren [8], the success of visualizations is determined by a user's ability to retrieve relevant information in an effective and efficient way.

The theory of cognitive fit, which heavily draws on information processing theory and cognitive load theory, has been used in various empirical investigations. To date,

recommendations pertaining to the optimal visualization design are contradicting, which is frequently attributed to the lack of knowledge of the visual perception process [1] [9,10,11,12].

Understanding how visualizations affect a user's perception is highly complex as it is influenced not only by the task and the data at hand but also by individual factors such as experience, knowledge or culture [13,14,15,16]. A generally accepted method for determining the quality of graphical representations a priori is still missing [15] [17]. Furthermore, user differences in personality and cognitive factors are crucial for their evaluation of visualizations [15]. In this paper we therefore introduce a model which tests antecedents of cognitive efficiency and accounts for the cognitive burden of various types of visualizations.

2 Theoretical Background

Our model is based on information processing theory and cognitive load theory. The former differentiates between three essential forms of memory: sensory store (reception of environmental information for a few seconds), short-term memory (analyzes, deconstructs and synthesizes information), and long-term memory (responsible for creating and saving mental constructs or schemas) [18,19]. Cognitive load theory provides guidelines on how to foster information retrieval and learning. This can be achieved in two ways. Germane cognitive load, which is the load devoted to learning new cognitive schemas [20], can be enhanced through standardized or well-known formats. Extraneous cognitive load, which is the additional load placed on the user by the design or the task, can be improved with the right visualization formats and designs [21].

Cognitive fit theory indicates how the extraneous cognitive load can be enhanced. It states that visualization needs to fit the task at hand. If this is the case the problem solver does not need to exert additional cognitive effort to either transform the problem representation to better match the task or to transform decision processes to better match problem representation [22]. However, this theory produced contradicting results, indicating that individual influences need to be considered as moderating or mediating variables [1] [16,17].

We conducted an extensive literature review that resulted in a total of 1,952 articles based on a keyword search which was finally reduced to 237 articles which actually fit our research goal. Four essential dimensions of information visualization are prominent in the literature (see Table1).

Table 1: Four Dimensions of Information Visualization

Dimensions	Description
Visual Complexity	Visual complexity is the degree of difficulty to transform an image into a consistent verbal description [16]. Two components determine visual complexity, namely the *visualization type* and the *design* or *structure* of a given visualization [7] [23,24].
Task Complexity	Task complexity has three determinants: *task type*, *task difficulty* (sometimes also referred to as complexity) and *task environment* [25]. According to Vessey [22] task type includes spatial (i.e. relationships and comparison making) and symbolic tasks

	(i.e. usage of discrete data values). Hard and Vanecek [26] highlight the importance of accumulation (i.e. the acquisition and recall of a single information cue), recognition (i.e. the recognition of patterns and relationships between 2-3 information cues), estimation (i.e. the identification of trends between numerous information cues), and projection (i.e. prediction of future values). Task difficulty can either be calculated objectively [27] or tested subjectively by asking participants. Finally, task environment represents external factors such as time constraints and task interruptions [28].
Data Complexity	Data complexity combines *data type* and *data density*. Data type comprises dimensions [24] and data density accounts for the amount of data compressed in a visualization [29].
Individual Complexity	Individual complexity can be clustered into three dimensions: Cognitive traits represent a person's working memory ability, cognitive states represent situational and emotional influences as well as experience and biases [15]. Situational and emotional states refer to a persons' origin and current state (e.g. being tired or depressed). Other important factors include knowledge and expertise, experience, decision making style, and gender as well as motivation, concentration and emotional issues that might focus attention [6] [19] [21] [30].

3 Hypotheses Development

In this paper we investigate the quality of visualization for managerial decision making. The focus lies on the examination of some of the most frequently mentioned individual influences in literature. According to cognitive fit theory, individual differences impact perception, but there is a lack of empirical investigation [1] [16,17].

Two of the most frequently mentioned influences on information processing efficiency are prior knowledge (i.e. experience) [1] [7] and domain expertise.

H1: A higher level of domain expertise positively influences efficiency in cognition.

H2: A higher level of experience positively influences efficiency in cognition.

Furthermore, spatial ability (sometimes also referred to as working memory capacity or perceptual speed) is said to influence the time needed to process visual information [1] [8] [17]. The higher a person's spatial ability, the faster the processing of information processing will take place.

H3: A higher level of spatial ability positively influences efficiency in cognition.

With respect to cultural background two of Hofstede's cultural dimensions are associated with information processing and information overload: uncertainty avoidance and long term orientation [31].

H4: A lower "Uncertainty Avoidance Index" positively influences efficiency in cognition.

H5: A higher "Long Term Orientation Index" positively influences efficiency in cognition.

4 Research Methodology

We conducted a laboratory experiment in which subjects were given four different tasks while viewing 18 different visualizations, which included different versions of bar charts, column charts, and tables. All of these visualizations showed a company's financial performance and all of the tasks had an optimal solution. Participants were selected from a student population with at least one year background in business administration. During a 30 minute session, tasks were given to subjects by a computer-based decision support system and eye tracking data was recorded. In total, 84 students volunteered to participate in the experiment which resulted in a total of 1,476 data records, since each student completed multiple tasks. Participants were randomly assigned to one of four experimental groups. Using various parametric and nonparametric tests, we found no significant differences across treatments according to gender, age, years in school, major, and prior experience.

A 4x9x2 between-subjects and within-subjects experimental design was used with four levels of task type based on Vessey [22] and Hard and Vanecek [26] (*accumulation*, *recognition*, *estimation*, *projection*), nine levels of information presentation format, and two levels of information types (i.e. time series data and data on the structural split of e.g. revenue within one year on the product mix of the company). Randomization of the tasks within the four groups was used.

Our dependent variable was efficiency. Participants were able to determine the pace of the experiment by independently going through the test by clicking. No time constraints were imposed. Efficiency was measured with the time span between seeing the visualization and stating the answer (net dwell time) and by the total fixation count per participant for each task. The latter was recorded using an eye tracking device (SMI RED with a sampling rate of 120 Hz, a nine point calibration and a 4 point validation). When analyzing eye tracking, data fixations are of particular interest. They are short stops where the eye is able to process information. No information can be processed during a movement of the eye, which is called saccade. Longer fixations are associated with greater visual and/or cognitive complexity. An increased number of fixations can be interpreted as having a negative impact on search efficiency [32,33].

Additional factors used in the model include domain expertise, spatial ability, culture, task complexity and visual complexity. Expertise with charts, working experience, years of school education, and visual complexity were measured via self-reported data. Spatial ability was measured by operating span and symmetry span which were collected with an automated test using E-Prime, a software frequently used for psychological tests [34]. Cultural dimensions, such as Uncertainty Avoidance and Long Term Orientation, were measured using Hofstede's scales [35] and task complexity was calculated according to Wood [27].

5 Results

We used PLS modeling to test the five hypotheses within a larger context. For modeling we used SmartPLS [36] and calculated the total effects (direct and indirect) of the path coefficients (see Table 2).

Table 2. Path Coefficients Statistics

	Sample	T-Statistics	P-Values	Hypothesis	
DomainExpertise -> Efficiency	0.071	2.563	0.011	H1	✓
Experience -> Efficiency	0.116	1.003	0.316	H2	✗
SpatialAbility -> Efficiency	0.059	2.275	0.023	H3	✓
Uncertainty Avoidance Index -> Efficiency	0.072	1.917	0.056	H4	~
Longterm Orientation Index -> Efficiency	-0.096	3.006	0.003	H5	✓

In line with previous academic studies, individual influences do have an influence on information processing efficiency, however, not all of them show significant results. As can be seen in Table 2 domain expertise, spatial ability, and long term orientation do have a significant effect on efficiency, while experience and uncertainty avoidance turn out to be not significant.

6 Discussion and conclusion

In this study we found that individual characteristics significantly impact information processing efficiency. Three of the five tested influential factors showed significant results, namely domain expertise, spatial ability, and long term orientation. Our preliminary results provide evidence that the neglect of individual factors might be the reason for producing contradicting results when determining adequate visualizations a priori. Knowing the importance of those characteristics might help researchers and companies using visualizations to better target their audience, which will in turn increase the capacity of information perception and improve decision making.

In this study subjects are rather homogeneous and these results need to be further investigated using a more heterogeneous sample. In follow-up studies we will therefore gather data from different regions to determine whether individual cultural differences do exert an influence on information visualization. Additionally, other factors need to be investigated such as the motivation of the participants [5], decision-making ability [31], personality style [16], and other cultural dimensions based on Hofstede [17].

7 References

1. Dilla, W., Janvrin, D.J., Raschke, R.: Interactive Data Visualization: New Directions for Accounting Information Systems Research. Journal of Information Systems 24(2), 1-37 (2010)
2. Edmunds, A., Morris, A.: The Problem of Information Overload in Business Organizations: a Review of the Literature. International Journal of Information Management 20(1), 17-28 (2000)
3. Eppler, M.J, Mengis, J.: The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. Information Society 20, 35-344 (2004)

4. Lurie, N.H., Mason, C.H.: Visual Representation: Implications for Decision Making. Journal of Marketing 71(1), 160-177 (2007)
5. Tortosa-Edo, V., López-Navarro, M.A., Llorens-Monzonís, J., Rodríguez-Artola, R.M.: The Antecendent Role of Personal Environment Values in the Relationships among Trust in Companies, Information Processing and Risk Perception. Journal of Risk Research 17(8), 1019-1035 (2014)
6. Al-Kassab, J., Ouertani, Z.M., Schiuma, G., Neely, A.: Information Visualization to Support Management Decisions. International Journal of Information Technology & Decision Making 13(2), 407-428 (2014)
7. Parsons, P., Sedig, K.: Adjustable Properties of Visual Representations: Improving the Quality of Human-information Interaction. Journal of the Association for Information Science & Technology 65(3), 455-482 (2014)
8. Conati, C., Maclaren, H.: Exploring the Role of Individual Differences in Information Visualization. In: Proceedings of AVI 2008, pp. 199-206. ACM (2008)
9. Barat, A. H.: Human Perception and Knowledge Organization: Visual Imagery. Library Hi Tech 25(3), 338-351 (2007)
10. Galletta, D., Vessey, I.: Cognitive Fit: An Empirical Study of Information Acquisition. Information Systems Research 2(1), 63-84 (1991)
11. Porat, T., Oron-Gilad, T., Meyer, J.: Task-dependent Processing of Tables and Graphs. Behaviour & Information Technology 28(3), 293-307 (2009)
12. So, S., Smith, M.: Multivariate Decision Accuracy and the Presentation of Accounting Information. Accounting Forum 28(3), 283-305 (2004)
13. Dilla, W.N., Janvrin, D.J.: Voluntary Disclosure in Annual Reports: The Association Between Magnitude and Direction of Change in Corporate Financial Performance and Graph use. Accounting Horizons 24(2), 257-278 (2010)
14. Parush, A., Hod, A., Shtub, A.: Impact of Visualization Type and Contextual Factors on Performance with Enterprise Resource Planning Systems. Computers & Industrial Engineering 52(1), 133-142 (2007)
15. Peck, E.M., Yuksel, B.R., Harrison, L., Ottley, A., Remco, C.: Position Paper: Towards a 3-dimensional Model of Individual Cognitive Differences. In: Proceedings of BELIV 2012, pp. 1-6. ACM (2012)
16. Yigitbasioglu, O.M., Valcu, O.: A Review of Dashboards in Performance Management: Implications for Design and Research. International Journal of Accounting Information Systems 13(1), 41-59 (2012)
17. Ziemkiewicz, C, Kosara, R.: Beyond Bertin: Seeing the Forest despite the Trees. IEEE Computer Graphics & Application 30(5), 7-11 (2010)
18. Miller, G.A.: The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review 63, 81-97 (1956)
19. Lord, R.G., Maher, K.J.: Alternative Information-processing Models and their Implications for Theory, Research, and Practice. Academy of Management Review 15(1), 9-28 (1990)
20. Anderson, E.W., Potter, K.C., Matzen, L.E., Shepherd, J.F., Preston, G.A., Silva, C.T.: A User Study of Visualization Effectiveness Using EEG and Cognitive Load. Eurographics / IEEE Symposium on Visualization 2011 30(3), 791-800 (2011)
21. Mostyn, G.R.: Cognitive Load Theory: What It Is, Why It's Important for Accounting Onstruction and Research. Issues in Accounting Education 27(1), 227-245 (2012)
22. Vessey, I.: Cognitive Fit: A Theory-based Analysis of the Graphs versus Tables Literature. Decision Science 22(2), 219-240 (1991)
23. Hill, W.Y., Milner, M.M.: Guidelines for Graphical Displays in Financial Reporting. Accounting Education 12(2), 135-157 (2003)

24. Kuang, X., Zhang, H., Zhao, S., McGuffin, M.J.: Tracing Tuples across Dimensions: A Ccomparison of Scatterplots and parallel Coordinate Plots. Computer Graphics Forum 31(3), 1365-1374 (2012)
25. Speier, C., Vessey, I., Valacich, J.S.: The Effects of Interruptions, Task Complexity, and Computer-supported Decision-making Performance. Decision Sciences 34(4), 771-797 (2003)
26. Hard, N.J., Vanecek, M.T.: The Implications of Task and Format on the Use of Financial Information. Journal of Information Systems 5(2), 35-49 (1991)
27. Wood, R.E.: Task Complexity: Definition of the Construct. Organizational Behavior and Human Decision Processes 37, 60-82 (1986)
28. Speier, C.: The Influence of Information Presentation Formats on Complex Task Decision-making Performance. International Journal of Human-Computer Studies 64(11), 1115-1131 (2006)
29. Gelman, A., Unwin, A.: InfoVis and Statistical Graphics: Different Goals, Different Looks. Journal of Computational & Graphical Statistics 22(1), 2-28 (2013)
30. Hahn, U.: Experiential Limitation in Judgment and Decision. Topics in Cognitive Science 6(2), 229-244 (2014)
31. Kook, N., Parente, R., Verville, J.: Can Hofstede's Model Explain National Differences in Perceived Information Overload? A Look at Data from the US and New Zealand. IEEE Transactions on Professional Communication 51(1), 33-49 (2008)
32. Goldberg, J., Helfman, J.: Eye Tracking on Visualizations: Progressive Extraction of Scanning Strategies. In: Huang, W. (ed.) Handbook of Human Centric Visualization, pp. 337-372. Springer (2014)
33. Renshaw, J.A., Finlay, J.E., Tyfa, D., Ward, R.D.: Designing for Visual Influence: An Eye Tracking Study of the Usability of Graphical Management Information. In: Proceedings of INTERACT 20003), pp. 144-151. ACM (2003)
34. Redick, T.S., Broadway, J.M., Meier, M.E., Kuriakose, N.U., Kane, M.J., Engle, R.W.: Measuring Working Memory Capacity with Automated Complex Span Tasks. European Journal of Psychological Assessment 28 (3), 164-171 (2012)
35. Hofstede, G.: Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations - 2nd Edition. Sage Publications (2001)
36. Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M.: A Primer on Partial Least Squares Structural Equation ;odeling (PLS-SEM). Sage Publishing (2014)

Using fMRI to Explain the Effect of Dual-Task Interference on Security Behavior

Bonnie Brinton Anderson[1], Anthony Vance[1], Brock Kirwan[1], Jeffrey Jenkins[1], and David Eargle[2]

[1] Brigham Young University, Provo, Utah, USA
{bonnie_anderson,anthony.vance,kirwan,jeffrey_jenkins}@byu.edu
[2] University of Pittsburgh, Pittsburgh, Pennsylvania, USA
dave@daveeargle.com

Abstract. We examine how security behavior is affected by dual-task interference (DTI), a cognitive limitation in which even simple tasks cannot be simultaneously performed without significant performance loss. We find that security messages that interrupt users actually make users more vulnerable by increasing security message disregard—behaving against the recommended course of action of a security message. We study the previously unexamined effect of DTI on a secondary, interrupting task—a security message. In a security context, it is critical that his interruption be carefully heeded.

We use functional magnetic resonance imaging (fMRI) to explore (1) how DTI occurs in the brain in response to interruptive security messages and (2) how DTI influences security message disregard. We show that neural activation in the medial temporal lobe (MTL)—a brain region associated with declarative memory—is substantially reduced under a condition of high DTI, which in turn significantly predicts security message disregard.

Keywords: functional magnetic resonance imaging (fMRI) · dual-task interference (DTI) · security behavior.

1 Introduction and Theory

Security messages, which prompt the user to perform a security action, frequently interrupt users. Unfortunately, people often behave against the messages' recommended course of action—a behavior known as *security message disregard* [1]—notwithstanding their critical importance. We show that fundamental cause of security message disregard is dual-task interference (DTI), a limitation of the human cognitive system [2] in which the human brain processes tasks serially and, therefore, must rapidly switch attention between multiple tasks that are being attempted at the same time [3].

The objectives of this study are twofold. First, we aim to explore how DTI occurs in the brain in response to interruptive security messages. To do so, we take a *NeuroIS* approach—the application of neuroscience literature and methods to IS. Specifically,

we used functional magnetic resonance imaging (fMRI) to observe DTI as it occurs in the brain in response to a security message and a competing primary task. Additionally, we seek to explain how DTI in the brain affects security message disregard.

Under the divided attention paradigm of DTI [2], participants must switch attention between stimuli. In our context, this includes switching attention when a security message interrupts a primary task. DTI occurs as individuals' cognitive functions are still engaged in the primary task while they are responding to the security message. Research [4] suggest that interruptive security warnings are often ignored or suboptimally addressed because users have a limited cognitive ability to switch between tasks.

Although DTI has been examined in a variety of contexts [3], how it may influence security message disregard is still unknown. Responses to security messages are likely to be especially susceptible to DTI because they are typically *secondary tasks* that interrupt the completion of a concurrent *primary task* that the user originally intended to accomplish. Further, the context of security messages is also unique because, whereas performance of the primary task is typically considered more important than the interrupting secondary task, carefully attending and responding to the security message—the interruption itself—is critically important in a security context.

2 fMRI Experiment

We predict that DTI will influence activation in the MTL, the brain region associated with long-term declarative memory or memory for facts and events. When people learn how to behave securely this information is stored in the MTL-dependent declarative memory system. When responding to security messages, people must retrieve information from declarative memory to generate a proper response based on past training, experience, and other stored information. Recalling information from even very recent training to behave securely requires use of declarative memory [5].

Conditions of High-DTI (responding to a security message in the middle of another task) result in lower activation in the MTL associated with recalling security information than in the Warning-Only task. The brain often cannot meet the demands of the multiple tasks simultaneously. Thus, DTI inhibits one's ability to activate the MTL in response to the security message. Thus, we hypothesize:

H1. In the MTL region of the brain, activity will be lower under the High-DTI condition as compared to the Warning-Only condition.

Literature has extensively validated the relationship between DTI and task performance even in simple tasks [6]. Decreased activation in the MTL resulting from High-DTI will likely lead to users not being able to access information from declarative memory to evaluate security messages. Performance will thereby decrease, as security behavior will have been informed possibly by inadequate information and processing.

H2. Security message disregard will be higher under the High-DTI condition as compared to the Warning-Only condition.

Building on our previous hypotheses, we hypothesize that the difference in MTL activation between High-DTI and Warning-Only tasks should predict the change in security message disregard between the two conditions. In summary,

H3. For the MTL region of the brain, the change in activation between the High-DTI and Warning-Only conditions will positively predict the change in security message disregard between the High-DTI and Warning-Only conditions.

2.1 Methodology

We used a repeated-measure, within-subject experimental design that required participants to respond to security warnings that either interrupted or did not interrupt a primary task. The security messages used in this experiment were operationalized as permission warnings similar to those that are displayed as users install a Google Chrome browser extension. Prior to beginning the treatments, participants received training regarding acceptable and risky permissions. Security message disregard was measured as inappropriate installation of a risky extension. Participants completed the experiment in both treatments presented in a random order.

In the High-DTI treatment, participants were presented with a seven-digit number. They were asked to encode the number for 5 seconds. Then the number disappeared and a warning was shown during the rehearsal phase of the recall task. Participants were given 7 seconds with a jitter of +/-3 seconds to click on either reject or accept based on their previous training. Next, a question appeared asking participants to select the number they were most recently asked to memorize among five other numbers. Participants were given 7 seconds to select the number, and then given a break for 7 seconds (+/-3) to be used as a baseline in the analysis. Participants repeated this 18 times.

In Warning-Only treatment, participants only evaluated warnings and did not receive the encode/retrieve task. Like the previous treatments, participants were given 7 (+/-3) seconds to respond to the warning. This was repeated 18 times with a break between each trial to be used as a baseline in the analysis.

After a successful pilot test using Amazon's Mechanical Turk service, we ran the full experiment in an fMRI laboratory to investigate the neural correlates of DTI in our chosen tasks. Participants were verbally informed about the MRI procedures and the experimental task. Participants viewed the experimental images on a large MR-compatible monitor at the opening of the MRI scanner by means of a mirror attached to the head coil. Participants used a trackball to interact with the security warnings and memorization task throughout the experiment. We recruited 24 participants (13 male, average age 23.7 years) from the university community.

2.2 Analysis

We examined the neural correlates of responding to security warnings under dual-task conditions by comparing activation for the High-DTI warning/rehearsal period with activation for the warning in the Warning-Only condition using paired *t*-tests. We exclusively masked the results of this comparison with the warning versus baseline We found activation was greater in the MTL for the Warning-Only condition than for

the High-DTI condition ($t(23) = 3.534$, $p < .005$), suggesting that participants were utilizing the MTL more for processing the security warning in the Warning-Only condition, supporting H1 (see Figure 1).

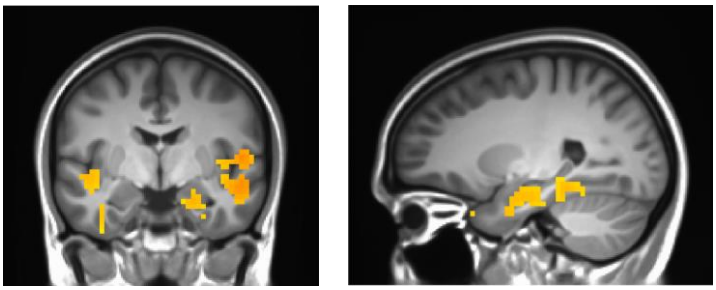

Fig 1. Decreased activity in response to the High-DTI compared to the Warning-Only

In addition to the fMRI analysis, we explored how DTI influenced participants' actual security message disregard. Security message disregard was significantly higher (22.92% vs. 7.41%) in the High-DTI treatment than in the Warning-Only treatment ($\chi 2(1) = 40.391$ $p < .01$), supporting H2.

We next explored whether the change in MTL activation between the High-DTI and Warning-Only treatments predicts participants' change regarding security message disregard. We specified a regression model with participants' change in terms of security message disregard as the dependent variable and participants' change in MTL activation between the two treatments as the independent variable. The results support the notion that the change in MTL activation significantly influences security message disregard: $\beta = -0.519$, $t(23) = 2.844$, $p < .01$, $R^2 = 0.269$, supporting H3.

We found that participants in the High-DTI treatment exhibited less activation in the bilateral MTL than participants in the Warning-Only treatment. This suggests that DTI inhibits one's ability to utilize the MTL to retrieve information from the long-term memory necessary to accept or reject the permission warnings. As such, we found that people had more than 15% higher security message disregard in the High-DTI treatment than in the Warning-Only treatment. We found that the change in MTL predicted participants' change in terms of warning response accuracy.

We also tied the fMRI and behavioral performance data by showing that decreases in MTL activation under a condition of High-DTI directly predict participants' increased security message disregard. Thus, our linkage of fMRI and behavioral data provides strong evidence of the influence of DTI on security message disregard.

3 Contributions

This paper demonstrates that DTI suppresses activity in the MTL region of the brain, which decreases one's ability to retrieve the necessary information from declarative memory to properly respond to the security message. This provides a sound theoretical foundation for objectively measuring the influences of DTI in the brain for security messages and other system-generated alerts, and also for designing messages that better engage the MTL to improve security behavior.

Second, our research shows that the change in activation in the MTL regions of the brain predicts security message disregard. A regression analysis indicated that the change in MTL activation between treatments alone accounted for 26.9% of the variance in security message disregard behavior in one analysis and 22.1% in the other. Thus, we contribute by directly tying fMRI data and behavioral performance data, providing a powerful objective predictor of security message disregard.

4 Conclusion

Users frequently fail to appropriately respond to security messages. In this paper, we explore a limitation of the human brain that contributes to security message disregard—dual task interference (DTI). In previous literature, DTI is primarily used to explain how performance decreases in primary tasks (e.g., a work-related task) when a secondary task interrupts or is performed concurrently. In our research, we use DTI theory and neuroscience to explain how the secondary task—in our context, the security message—may also experience decreased performance when it interrupts a primary task. Using functional magnetic resonance imaging (fMRI), we examine how DTI occurs in the brain for security warnings. When a security warning interrupts another primary task, activation decreases in the medial temporal lobe (MTL)—a brain region used to recall relevant security information from declarative memory. Consequently, security message disregard also decreases.

Acknowledgements. This research was funded by NSF Grants CNS-1422831 and DGE-1247842.

References

1. Vance, A., Anderson, B.B., Kirwan, C.B., Eargle, D.: Using Measures of Risk Perception to Predict Information Security Behavior: Insights from Electroencephalography (EEG). Journal of the Association for Information Systems 15(10), 679-722 (2014)
2. Szameitat, A.J., Schubert, T., Muller, H.J.: How to Test for Dual-task-specific Effects in Brain Imaging Studies: An Evaluation of Potential Analysis Methods. NeuroImage 54(3), 1765-1773 (2011)
3. Pashler, H.: Dual-task Interference in Simple Tasks: Data and Theory. Psychological Bulletin 116(2), 220-244 (1994)
4. Bravo-Lillo, C., Cranor, L.F., Downs, J., Komanduri, S., Sleeper, M.: Improving Computer Security Dialogs. In: Proceedings of INTERACT 2011, Part IV, LNCS 6949, pp. 18-35. Springer, Berlin, Heidelberg (2011)
5. Friedman, H., Goldman-Rakic, P.: Activation of the Hippocampus and Dentate Gyrus by Working-memory: A 2- Deoxyglucose Study of Behaving Rhesus Monkeys. Journal of Neuroscience 8(12), 4693-4706 (1988)
6. Borst, J.P., Taatgen, N.A., Van Rijn, H.: What Makes Interruptions Disruptive? A process-model Account of the Effects of the Problem State Bottleneck on Task Interruption and Resumption. In: Proceedings of CHI 2015. ACM (2015)

Measuring Appeal in Human Computer Interaction: A Cognitive Neuroscience-Based Approach

Tillmann Neben[1,*], Bo Sophia Xiao[2], Erik Lim[3], Chee-Wee Tan[4], and Armin Heinzl[1]

[1] University of Mannheim, Germany
{neben,heinzl}@uni-mannheim.de
[2] University of Hawai'i at Mānoa, USA
boxiao@hawaii.edu
[3] University of New South Wales, Australia
e.t.lim@unsw.edu.au
[4] Copenhagen Business School, Denmark
cta.itm@cbs.dk

Abstract. Appeal refers to the positive emotional response to an aesthetic, beautiful, or in another way desirable stimulus. It is a recurring topic in information systems (IS) research, and is important for understanding many phenomena of user behavior and decision-making. While past IS research on appeal has relied predominantly on subjective self-rating scales, this research-in-progress paper proposes complementary objective measurement for appeal. We start by reviewing the linkages between the theoretical constructs related to appeal and their neurophysiological correlates. We then review past approaches to measuring appeal and discuss their characteristics. Finally, we arrive at a recommendation that builds on a combination of psychophysiology (EDA, facial EMG) and brain imaging (fNIRS).

1 Introduction

A recurring topic in information systems (IS) research, appeal refers to the positive emotional response to an aesthetic, beautiful, or in another way desirable stimulus. IS scholars have incorporated this construct in studying various phenomena such as technology acceptance and use [1], issues related to eCommerce and trust [2], website design [3,4], and culture [5]. The majority of existing studies rely on self-rating scales for assessing the subjective perception of appeal. The limitations and fallacies of assessing short-lived and partly unconscious emotional responses with self-rating scales have been discussed in prior literature [6-9]. In this paper, we aim to address this gap in previous research by identifying and explaining neuroscience methods for measuring appeal. Extant literature indicates that the psychological states of appeal, pleasantness, attractiveness, and charm are closely linked, and that they are important in determining online behavior and decision-making. We hence include these constructs in our investigation, and follow the reasoning of [4, 10] by treating them as coherent constructs under the umbrella term of *appeal*.

2 Relevance and measurement of appeal-related emotional states

Over the last few years, there has been a stream of IS research on appeal that focuses on how website design fosters *visual appeal*, which in turn reinforces consumers' positive beliefs and attitudes toward the website. For instance, [3] found that human images on an eCommerce website generated image appeal, which positively influenced consumers' trust in the website. [11] as well as [12] further probed how the color scheme of a website affected color appeal, which in turn increased consumers' trust in, satisfaction with, and intention to purchase from the website. Finally, in their research within an eCommerce context, [4], [13], and [14] empirically validated visual appeal as an critical antecedent of website aesthetics, perceived relationship rewards, and website quality. Insomuch as a large part of website and product appeal is driven by visual cues, we focus on *visual appeal* in the current investigation. Though appeal can involve cognition, it is primarily described as an emotional response [15]. Measuring an emotional reaction to a stimulus poses several challenges. Aesthetic emotion is expected to be a relatively subtle affective state compared to e.g. experiences episodes of strong technostress [16]. The following section reviews methods that satisfy these conditions and the neural correlates standing in relation to experiencing appeal.

2.1 Review of neural correlates of appeal

The theoretical constructs relating to appeal – and in particular to visual appeal – overlap significantly. The perception of visual aesthetic, beauty, pleasantness, and attractiveness all relate to visual appeal [4, 10]. Interestingly, similar overlaps are evident on the level of neural correlates termed "hedonic brain activation" [17-19]. The neural correlates of *aesthetics* as a broader term have been studied in many contexts, in relation to auditory stimuli [15] or product package design [20]. Studies using magnetencephalography (MEG) identified increased activity in the prefrontal cortex [21] as well as characteristic time-frequency changes [22] as correlates of aesthetic perception. These and related research works belong to a specialized community called Neuroaesthetics [23]. Three major brain imaging correlates of positive aesthetic experiences have been identified [22]: People engaged with aesthetically appealing stimuli show increased somatosensory cortical processing [24], and elicit increased activity in cortical regions involved with evaluative judgment [19,21]. They also show an activation of the cortical and subcortical brain regions that belong to the reward circuit (orbitofrontal cortex, striatum, among others) [25-28]. Whilst the first two correlates are linked to visually perceiving and mentally judging aesthetics, the third correlate appears to be more closely related to the emotional outcome of appeal perception. Indeed, [29] describes the ventral striatum as the hedonic "hot spot" of the brain, and [17] presents evidence that positive aesthetics processing relates to the "reward circuit" (orbitofrontal cortex, medial orbitofrontal cortex, striatum). These findings suggest that the evaluation and perception of likability and pleasure are au-

tomatic process that occur in orbitofrontal cortex areas and the ventral striatum [17]. Importantly, this medial orbitofrontal cortex (mOFC) and ventral striatal activation occurs for real as well as for hypothetical experiences [30], such as when looking at a desirable product online. In Table 1 we review the central correlates of appeal, beauty, and pleasantness, as discussed in the literature.

2.2 Review of neuroscience methods for measuring appeal

Next, we review existing neuroscience methods for approaching these correlates **Peripheral physiology** does not produce functional images of the brain areas discussed above. However, it allows the measurement of activations of the peripheral nervous system, and more specifically of the autonomous nervous systems (ANS). Past research, which relied on heart-rate variability (HRV) and electro-dermal activity (EDA) for assessment, has produced evidence that peripheral physiology can measure characteristic ANS reactions to aesthetic experiences [1,15,31,32]. Related to this, facial electromyography (fEMG) involves placing electrodes over specific muscle groups on the face. Minimal face muscle activation allows inferring affective states [33,34,35,36,37]. Previous studies have used fEMG for measuring pleasantness and appeal [38,39]. **Electroencephalograph** (EEG) measures electrical brain activity using sensors on the scalp. Prior research has utilized EEG to assess appeal and aesthetic perception [40]. **Functional magnetic resonance imaging** (fMRI) is a measure of brain activity. It uses the fact that more blood flows into more active brain areas (hemodynamics). Researchers have relied on fMIRI for assessing appeal and aesthetic perception. For example, Cupchick et al. used fMRI to reveal the role of bilateral insula in aesthetic perception [27]. **Functional near-infrared spectroscopy** (fNIRS) is a measure of brain activity that is related to fMRI. It uses the fact that skin, tissue, and bone material are almost transparent to specific spectra of infra-red light, but that hemoglobin (Hb) and deoxygenated-hemoglobin (deoxy-Hb) are not [41,42]. Researchers have relied on fNIRS for measuring appeal and aesthetic perception. For example, [18,43,44] used fNIRS in the context of aesthetics, beauty and positive reward perception.

2.3 Towards a measurement model of appeal in IS research

Next, we discuss these methods feasibility to support our research-in-progress on appeal and eCommerce. We emphasize their differences in validity and reliability as well as their temporal and spatial resolution. Our reasoning builds on past recommendations for method selection in NeuroIS [7]. We also discuss the costs associated with these methods as well as how extensive the necessary training is.

Psychophysiological measurements do not stand in a one-to-one relationship to mental processes. This causes a validity threat to many physiological methods. For example, EDA responds to positive affect but also to negative affect and to most other activations of the ANS [53]. Another issue is that most measurement of ANS activity is ultimately limited to its arousal component. That is, not the type of emotion is measured, but its strength.

Table 1. Review of most relevant correlates and methods

Ref.↓						
	Appeal (aesthetic, art)	EMG	EDA	fNIRS	EEG	fMRI
[15]	Skin conductance level and variability		✕			
[44]	Increase of oxygenated blood in the medial rostral prefrontal cortex during viewing of positive images			✕		
[45, 46]	Two-stage process with early anterior frontomedian activity after 300ms and right-hemisphere activity 600ms after stimuli				✕	
[10]	Increase in sensory (occipito-temporal) regions and the striatum. Also activation of the Default Mode Network					✕
[24]	Activation in bilateral occipital gyri, left cingulate sulcus, and bilateral fusiform gyri					✕
[8]	Frontomedial cortex, bilateral prefrontal and posterior cingulate, left temporal pole, and the temp. junction					✕
[20]	Activation in the nucleus accumbens and the ventromedial prefrontal cortex					✕
	Beauty (faces)					
[18]	Increased activity in the frontal and occipital cortexes			✕		
[47, 48]	Negativity 400ms over midline positions for *not beautiful* faces				✕	
[49]	Positive potential at about 30ms after stimulus onset (P300)				✕	
[25, 26]	Increase in orbito-frontal cortex activity					✕
[19]	Activation in frontomedian cortex, bilateral prefrontal and posterior cingulate, left temporal pole, and the temp. junction					✕
[50]	Activity in the medial orbito-frontal cortex (mOFC)					✕
	Pleasantness					
[38]	Zygomaticus major activity	✕				
[39]	Zygomaticus major activity, and incr. SCR	✕	✕			
[51]	Asymmetrical increase of theta and alpha activity in the left (right) hemisphere				✕	
[52]	Activity in the right anterior prefrontal cortex					✕
[17]	mOFC, ventral striatum (reward circuit)					✕

One solution is to combine EDA with fEMG. Whilst EDA provides a measure of arousal strength, fEMG provides a measure of emotional valence (positive vs. negative). Taken together, these data points can be mapped onto Russel's circumplex model of emotions [54]. This is supported by the finding of [55], who reports a strong arousal component in aesthetic perception, and past works that used fEMG for detecting positive valence emotions [33]. Psychophysiological measures are very reliable, as they measure ANS activity, which is by its definition autonomous and predictable. The temporal resolution of physiological measures is moderate. The cost for physiological tools is very low, compared to other tools (such as fMRI). Training effort is equally low: IS researchers can conduct first experiments with only a few months of setup. In summary, psychophysiology appears capable of measuring appeal, and physiological methods are comparatively easy to use. **Electroencephalography** studies have produced valid and reliable measures of brain activation that stands in relation to appeal. EEG measures the brain's electrical activity directly, whilst other brain imaging techniques, such as fNIRS and fMRI, measure changes in the blood flow. As a result, the temporal resolution of EEG is very high. However, and to a large part due

to reflections of the skull, the spatial resolution of EEG is lower than that of fNIRS or fMRI. The cost for EEG equipment and software is moderate. The training necessary for conducting a solid EEG study is extensive. In summary, EEG is capable of measuring appeal, but requires more extensive training and experience. **Functional magnetic resonance imaging** is a valid and very reliable measure of appeal-related constructs (cf. table 1). Compared to EEG, it has a poor temporal but good spatial resolution. However, the costs associated with fMRI are very high. This is true for the actual scanner hardware as well as for running the experiments. Additionally, most IS researchers will not find easy access to an fMRI. Moreover, the degree of training and skill for conducting fMRI experiments is very high. In most cases, IS researchers will need a trained neuroscientist in the team in order to conduct such a study. Another disadvantage is the unnatural position participants have to reside in, and the behavioral restriction the scanner tube imposes. In summary, fMRI is a highly sophisticated and capable brain imaging technology that is capable of measuring appeal-related constructs. However, it has practical limitations for those IS researchers who do not have an fMRI laboratory at their institution. **Functional near-infrared spectroscopy** is capable of measuring the correlates of emotional responses that occur in the reward circuit [56, 57]. Like fMRI, fNIRS has low temporal but good spatial resolution. The costs for fNIRS are moderate, and the requirements for training are moderate. In summary, fNIRS appears to be a capable method for measuring appeal and it is easier to operate and analyze when compared to other tools (such as fMRI).

Based on this and the findings of Table 1, it becomes evident that modern brain imaging technologies (such as fMRI) allow the most advanced measurements. The disadvantages of these methods, however, lie in the high costs and the extensive education and training required. Our review also shows that, while peripheral physiology allows less direct and less advanced measurements, these methods are comparatively cheap and easy to learn. Our review further reveals that fNIRS is positioned between these extremes. In conclusion, this research-in-progress proposes that if advanced methods such as fMRI are not available, a combination of psychophysiology (EDA and fEMG) and brain imaging (fNIRS) appears promising for measuring appeal. Table 2 recapitulates this proposed method configuration.

Table 2. The selected two methods this research-in-progress will pursue

	Psychophysiological	**fNIRS**
Correlate(s)	SCR, SCL (EDA), Zygomaticus major (EMG)	Reward circuit area
Temporal	Low	Less than EEG, but acceptable
Spatial	Indirect (via ANS)	High
Validity	Moderate	Good
Reliability	Good	Good
Practicality	Very low cost, moderate training, high external validity	Moderate cost, high training, very high external validity

3 Conclusion

In this paper, we review the correlates of appeal and identify the brain activation of the mOFC and ventral striatal as the most relevant brain sites [17, 30]. We also identify specific physiological responses that relate to appeal experiences. We then proceed to review the tools capable of measuring these correlates. Because extensive training requirements and concerns regarding the external validity speak against EEG and ERPs, and particularly fMIRI, we conclude that a promising way for measuring appeal is to rely on a combination of psychophysiological tools (EDA, fEMG) for peripheral measurement and fNIRS as a less intrusive and less costly but capable brain imaging method. The contribution of this research-in-progress work is three-fold: First, we provide a review of existing measures for appeal and related constructs. Second, we illustrate the advantages and disadvantages of prominent neuroscience tools when *subtle* and *short-lived* emotions are under investigation. Finally, we contribute by proposing a concrete method selection for measuring the experience of appeal.

4 References

1. Bhandari, U., Chang, K.: Role of Emotions and Aesthetics in ICT Usage for Underserved Communities: A NeuroIS Investigation. In: Proceedings of ICIS 2014. AIS (2014)
2. Pengnate, F., Sarathy, R.: Form and Function: How Website Characteristics Impact User Behavior. In: Proceedings of AMCIS 2011. AIS (2011)
3. Cyr, D., Head, M., Larios, H., Pan, B.: Exploring Human Images in Website Design: A Multi-Method Approach. MIS Quarterly 33, 539–566 (2009)
4. Cai, S., Xu, Y., Yu, J., De Souza, R.: Understanding Aesthetics Design for E-Commerce Web sites: a Cognitive-Affective Framework. In: Proceedings of PACIS 2008, Paper 60. AIS (2008)
5. Caishun, Z.: Cultural Recontextualization in E-business Web Design: A Socio-cognitive Model. In: Proceedings of WHICEB 2013, Paper 85. AIS (2013)
6. De Guinea, A.O., Titah, R., Léger, P.-M., Micheneau, T.: Neurophysiological Correlates of Information Systems Commonly Used Self-Reported Measures: A Multitrait Multimethod Study. In: Proceedings of 45th Hawaii International Conference on System Sciences (HICSS). AIS (2012)
7. Dimoka, A., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P.H., Pavlou, P.A., Müller-Putz, G., Riedl, R., Brocke, Vom, J., Weber, B.: On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. MIS Quarterly 36, 679-702 (2012)
8. Jacobsen, T.: Beauty and the Brain: Culture, History and Individual Differences in Aesthetic Appreciation. Journal of Anatomy 216, 184–191 (2010)
9. Sternberg, S.: Memory-scanning: Mental processes revealed by reaction-time experiments. American scientist 47, 421-457 (1969)
10. Vessel, E.A., Starr, G.G., Rubin, N.: The brain on art: intense aesthetic experience activates the default mode network. Frontiers in Human Neuroscience 6 (2012)
11. Cyr, D., Head, M., Larios, H.: Colour appeal in website design within and across cultures: A multi-method evaluation. International Journal of Human-Computer Studies 68, 1–21 (2010)

12. Pelet, J.-É., Papadopoulou, P.: The effect of colors of e-commerce websites on consumer mood, memorization and buying intention. European Journal of Information Systems 21, 438–467 (2012)
13. Campbell, D.E., Wells, J.D., Valacich, J.S.: Breaking the Ice in B2C Relationships: Understanding Pre-Adoption E-Commerce Attraction. Information Systems Research 24, 219–238 (2013)
14. Wells, J.D., Valacich, J.S., Hess, T.J.: What Signals Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions. MIS Quarterly. 35, 373–396 (2011)
15. Tschacher, W., Greenwood, S., Kirchberg, V., Wintzerith, S., van den Berg, K., Tröndle, M.: Physiological correlates of aesthetic perception of artworks in a museum. Psychology of Aesthetics, Creativity, and the Arts 6, 96–103 (2012)
16. Riedl, R., Kindermann, H., Auinger, A., Javor, A.: Technostress from a Neurobiological Perspective: System Breakdown Increases the Stress Hormone Cortisol in Computer Users. Business & Information Systems Engineering 4, 61–69 (2012)
17. Kühn, S., Gallinat, J.: The neural correlates of subjective pleasantness. NeuroImage 61, 289–294 (2012)
18. Fu, G., Mondloch, C.J., Ding, X.P., A Short, L., Sun, L., Lee, K.: The neural correlates of the face attractiveness aftereffect: A functional near-infrared spectroscopy (fNIRS) study. NeuroImage 85, 363–371 (2014)
19. Jacobsen, T., Schubotz, R.I., Höfel, L., Cramon, D.Y.V.: Brain correlates of aesthetic judgment of beauty. NeuroImage 29, 276–285 (2006)
20. Reimann, M., Zaichkowsky, J., Neuhaus, C., Bender, T., Weber, B.: Aesthetic package design: A behavioral, neural, and psychological investigation. Journal of Consumer Psychology 20, 431–441 (2010)
21. Cela-Conde, C.J., Marty, G., Maestú, F.: Activation of the prefrontal cortex in the human visual aesthetic perception. Proceedings of the National Academy of Sciences of the United States of America 101(16), 6321-6325 (2004)
22. Munar, E., Nadal, M., Castellanos, N.P.: Aesthetic appreciation: event-related field and time-frequency analyses. Frontiers in Human Neuroscience 5 (2011)
23. Chatterjee, A.: Neuroaesthetics: A Coming of Age Story. Journal of Cognitive Neuroscience 23, 53–62 (2011)
24. Vartanian, O., Goel, V.: Neuroanatomical correlates of aesthetic preference for paintings. Neuroreport 15, 893-7 (2004)
25. Kirk, U., Skov, M., Christensen, M.S., Nygaard, N.: Brain correlates of aesthetic expertise: A parametric fMRI study. Brain and Cognition 69, 306–315 (2009)
26. Kawabata, H.: Neural Correlates of Beauty. Journal of Neurophysiology 91, 1699–1705 (2004)
27. Cupchik, G.C., Vartanian, O., Crawley, A., Mikulis, D.J.: Viewing artworks: Contributions of cognitive control and perceptual facilitation to aesthetic experience. Brain and Cognition 70, 84–91 (2009)
28. Cela-Conde, C.J., Agnati, L., Huston, J.P., Mora, F., Nadal, M.: The neural foundations of aesthetic appreciation. Progress in Neurobiology 94, 39–48 (2011)
29. Pecina, S., Smith, K.S., Berridge, K.C.: Hedonic Hot Spots in the Brain. The Neuroscientist 12, 500–511 (2006)
30. Kang, M.J., Rangel, A., Camus, M., Camerer, C.F.: Hypothetical and Real Choice Differentially Activate Common Valuation Areas. Journal of Neuroscience 31, 461–468 (2011)
31. Christie, M.J.: Electrodermal activity in the 1980s: a review. Journal of the Royal Society of Medicine 74, 616 (1981)
32. Carbon, C.C., Michael, L., Leder, H.: Design evaluation by combination of repeated evaluation technique and measurement of electrodermal activity. Research in Engineer-

ing Design 19, 143-149 (2008)
33. Cacioppo, J.T., Petty, R.E., Losch, M.E., Kim, H.S.: Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. Journal of Personality and Social Psychology 50, 260-8 (1986)
34. Winkielman, P., Cacioppo, J.T.: Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. Journal of Personality and Social Psychology 81, 989–1000 (2001)
35. van Boxtel, A.: Facial EMG as a tool for inferring affective states. In: Proceedings of Measuring Behavior 2010, pp. 104–108. Eindhoven, The Netherlands (2010)
36. Larsen, J.T., Norris, J.I.: A facial electromyographic investigation of affective contrast. Psychophysiology 46, 831–842 (2009)
37. Neta, M., Norris, C.J., Whalen, P.J.: Corrugator muscle responses are associated with individual differences in positivity-negativity bias. Emotion 9, 640–648 (2009)
38. ZHOU, R., HU, S.: Effects of Viewing Pleasant and Unpleasant Photographs on Facial EMG Asymmetry. Perceptual and Motor Skills 99, 1157–1167 (2004)
39. Soussignan, R., Schaal, B., Rigaud, D., Royet, J.-P., Jiang, T.: Hedonic reactivity to visual and olfactory cues: Rapid facial electromyographic reactions are altered in anorexia nervosa. Biological Psychology 86, 265–272 (2011)
40. De Tommaso, M., Pecoraro, C., Sardaro, M., Serpino, C., Lancioni, G., Livrea, P.: Influence of aesthetic perception on visual event-related potentials. Consciousness and Cognition 17, 933–945 (2008)
41. Scholkmann, F., Wolf, M.: Measuring brain activity using functional near infrared spectroscopy: a short review. Spectroscopy Europe, August 22 (2012)
42. Jobsis, F.F.: Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198, 1264–1267 (1977)
43. Misawa, T., Shimokawa, T., Hirobayashi, S.: Possibility for Predicting the Evaluation of Product Price in the Prefrontal Cortex: A NIRS Study. International Journal of Engineering and Innovative Technology 4, (2014)
44. Kreplin, U., Fairclough, S.H.: Activation of the rostromedial prefrontal cortex during the experience of positive emotion in the context of esthetic experience. An fNIRS study. Frontiers in Human Neuroscience (2013)
45. Jacobsen, T., Hoefel, L.: Descriptive and Evaluative Judgment Processes: An Event-related Brain Potential Analysis of Processing Symmetry and Aesthetics. Journal of Cognitive Neuroscience (2000)
46. Höfel, L., Jacobsen, T.: Electrophysiological indices of processing aesthetics: Spontaneous or intentional processes? International Journal of Psychophysiology 65, 20–31 (2007)
47. Roye, A., Höfel, L., Jacobsen, T.: Aesthetics of Faces. Journal of Psychophysiology 22, 41–57 (2008)
48. Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G.: Electrophysiological Studies of Face Perception in Humans. Journal of Cognitive Neuroscience 8, 551–565 (1996)
49. Oliver-Rodriguez, J.C., Guan, Z., Johnston, V.S.: Gender differences in late positive components evoked by human faces. Psychophysiology 36, 176–185 (1999)
50. Ishizu, T., Zeki, S.: Toward a brain-based theory of beauty. PLoS One (2011)
51. Vecchiato, G., Toppi, J., Astolfi, L., De Vico Fallani, F., Cincotti, F., Mattia, D., Bez, F., Babiloni, F.: Spectral EEG frontal asymmetries correlate with the experienced pleasantness of TV commercial advertisements. Medical & Biological Engineering & Computing 49, 579–583 (2011)
52. Thakral, P.P., Moo, L.R., Slotnick, S.D.: A neural mechanism for aesthetic experience. Neuroreport 23, 310–313 (2012)
53. Figner, B., Murphy, R.O.: Using skin conductance in judgment and decision making research. In: Schulte-Mecklenbeck, M., Kuehberger, A., Ranyard, R. (eds.) A handbook

of process tracing methods for decision research, pp. 163–184. Psychology Press, New York (2011)
54. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39, 1161–1178 (1980)
55. Berlyne, D.E.E.: Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation. Hemisphere Publishing, Washington (1974)
56. Perlman, S.B., Luna, B., Hein, T.C., Huppert, T.J.: fNIRS evidence of prefrontal regulation of frustration in early childhood. NeuroImage 85, 326–334 (2014)
57. Suda, M., Uehara, T., Fukuda, M., Sato, T., Kameyama, M., Mikuni, M.: Dieting tendency and eating behavior problems in eating disorder correlate with right frontotemporal and left orbitofrontal cortex: A near-infrared spectroscopy study. Journal of Psychiatric Research 44, 547–555 (2010)

Mobile App Preferences: What Role Does Aesthetics and Emotions Play?

Upasna Bhandari[1,*], Tillmann Neben[2], and Klarissa T. T. Chang[1]

[1] Department of Information Systems, School of Computing, National University of Singapore, Singapore
a0106246@u.nus.edu, changtt@comp.nus.edu.sg
[2] Chair of General Management and Information Systems, University of Mannheim, Germany
neben@uni-mannheim.de

Abstract. This research-in-progress reports on the development of a NeuroIS measurement model for studying the role of emotions in non-instrumental preferences. We aim at exploring the effects of emotions and aesthetics on users' preferences for mobile application. The context of mobile apps is interesting because the phenomenon of high initial adoption but very low retention is still unexplained. For this, we aesthetically manipulated mobile apps, and measured subjects' affective responses. Our approach builds on galvanic skin response (GSR) and surface electromyography of the face.

Keywords: Aesthetics · Emotions · Physiology

1 Introduction

The success story of products like iMac, Macbook and iPad explains the growing emphasis on aesthetics and style rather than usability and reliability [1]. Consequently the mobile applications (apps) market has also grown at an enormous rate and it can be argued that *"aesthetics might be the only way to make your product stand out"* if you want a chance at being used [1]. Aesthetics play an important role in understanding why people choose certain products over others. Users have been found to rapidly judge the aesthetics of a webpage reliably within 50 milliseconds [2].

Another issue in case of mobile apps is that initial adoption is high however retention rate is extremely low. It has also been reported that 26% of people who download apps only use them once. Vilnai and Rafaeli in their study pointed out that the switching costs for consumer is little in the growing age of e-commerce [3]. Thus getting first impressions right becomes even more critical. Similar results were found where users made judgments about visual stimulus in a very short amount of time [4]. Thus there is a need to study measures like mobile app preferences and their relationship with aesthetics to understand consumer choice.

Another factor that is important to consider when understanding mobile app preferences is emotions. Norman has worked extensively on the connection between emo-

tion and aesthetics [5,6]. He explains a three step process that involves both cognitive and affective responses to aesthetic stimuli. However with apps, users are exposed to it for a very short amount of time. Robin and Holmes in their work showed that users made a decision about the credibility of a website within 3.42 seconds [7]. Thus in this work we focus on Type 1 or the visceral processing of stimuli response. This is the most basic level of response to stimuli. It is instinctive and happens within a few second of exposure [6]. Lynch mentions in his study that users create a predisposed notion to find an attractive design as usable and the effect lingers long after the conscious, behavioral and reflective processing [8].

2 Theoretical Development

Contemporary research on emotion and user experience suggests that aesthetically pleasing objects affect our emotions positively [9]. Porat and Tractisnky found that aesthetics influenced consumers emotional states and consequently attitudes towards web stores [3], [10].

When consumers are making purchasing decision regarding mobile apps, they are inevitably influenced the most by 'look and feel' of the app that can elicit different emotions. However rarely has the effect of aesthetics on emotional responses been explored in case of mobile apps. Both arousal and valence dimensions have been known to contribute to decision making. Previous work has shown that classical aesthetics are related to valence based demotions while expressive aesthetics are linked to arousal dimension of emotion [11,12,13].

H1(a): Interfaces with high classical aesthetics lead to positive valence emotions as compared to interfaces with low classical aesthetics.

H1(b): Interfaces with high expressive aesthetics lead to high arousal emotions as compared to interfaces with lower expressive aesthetics.

Both arousal and valence dimensions of emotion contribute to preference decision. More specifically positive valence has found to be associated with a greater possibility of being preferred by the user as compared to negative valence emotions. Similar findings for arousal based emotions have been found. A highly pleasure inducing experience is expected to have a stronger impact on app preferences.

H2 (a): Positive valence emotions lead to greater overall app preference as compared to negative valence emotions.

H2 (b): High arousal emotions lead to greater overall app preference as compared to low arousal emotions.

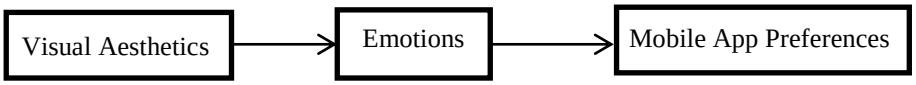

Fig. 1. Proposed Model to Understand Mobile App Preferences

3 Methodology

We intend to test the research model by conducting a lab experiment. The measurement of non-instrumental factors in IS so far has been restricted to traditional measures like surveys or self-report measures. Bodily physiological measures like EEG/MEG, fMRI etc. have only recently started to get attention [14]. They are not subjected to the fear of cognitive decision overshadowing the impact of affective processes. We intend to use electrodermal responses (EDA) for measuring arousal and facial electromyography (fEMG) for valence. For measuring mobile app preference, app choice is a behavioral measure for individual preference [15].

For creating aesthetically varied interfaces of the mobile app, we choose to manipulate three dimensions each from the classic and expressive scale as forwarded by [16]. We manipulate cleanliness, clarity and symmetry for classical aesthetics and creativity, originality and special effects for expressive aesthetics. We vary these six parameters to create four versions of the same app each with four pages each (home page, product page, payment page, checkout page): app 1 (low classic, low expressive), app 2 (low classic, high expressive), app 3 (high classic, low expressive) and app 4 (high classic and high expressive).

This study thus uses a 2 (classic aesthetics: high/low) * 2 (expressive: high/low) within subjects design. Due to space restrictions, we show three of the pages for app 4 in Fig 2. The reason for choosing within subject design is to garner enough statistical power and also because brain has been found to be idiosyncratic in nature and thus for our experiment design (we show all interfaces to all participants), within subject design is more effective. The participants will be assigned to each of the conditions randomly. While affective responses are collected unobtrusively via EDA and fEMG devices, a subjective question appears after each interface to collect the mobile application preference. A rest period of one minute is provided along with a filler task before next interface is shown. This is done to return the readings to baseline and avoid learning effect.

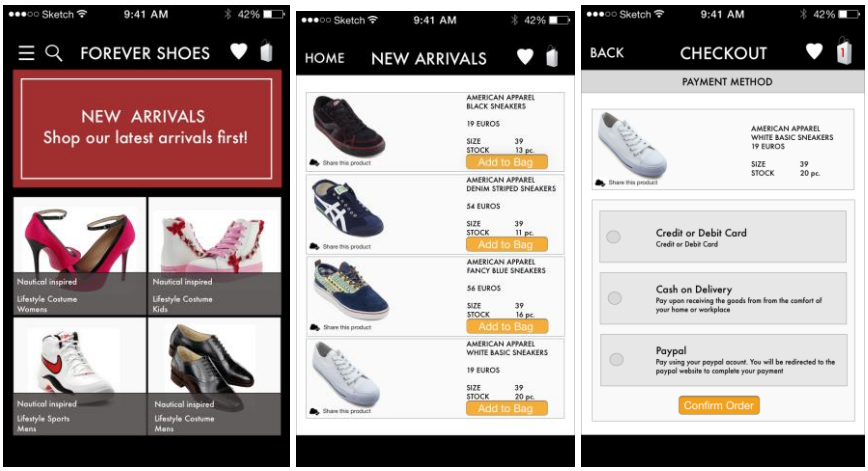

Fig. 2. Manipulated stimuli for high classic high expressive version (app 4)

4 Discussions and Conclusion

Our study will be able to demonstrate the role of aesthetics and help in understanding technology adoption from an emotional perspective. While usability and efficiency has been extensively explored in regards to technology acceptance, overall experiential impact of technology interaction has been overlooked [17]. Assuming that technology adoption is a purely cognitive decision is an underlying assumption with most frameworks addressing the phenomenon of technology use/adoption/continuance. However this can be insufficient where first impressions are concerned as they heavily rely on the emotional reactions of users [18]. We aim to demonstrate that in a unique case like mobile application where users can switch from one product to another at minimum cost, initial affective response to the aesthetic design becomes a crucial factor in deciding app preferences. Designers and managers can then use these insights into better designing their products so as to get the intended affective responses from users leading to higher app preferences.

References

1. Postrel, V.: Can Good Looks Really Guarantee a Product's Success? The New York Times, July 12, Section C, Page 2 (2001)
2. Lindgaard, G., Fernandes, G., Dudek, C., Brown, J.: Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression! Behaviour & Information Technology 25(2), 115-126 (2006)
3. Vilnai-Yavetz, I., Rafaeli, A.: Aesthetics and Professionalism of Virtual Servicescapes. Journal of Service Research 8(3), 245-259 (2006)
4. Alsudani, F., Casey, M.: The Effect of Aesthetics on Web Credibility. In: Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology, pp. 512-519. British Computer Society (2009)
5. Norman, D.: Emotion & Design: Attractive Things Work Better. Interactions 9(4), 36-42 (2002)
6. Norman, D.A.: Introduction to This Special Section on Beauty, Goodness, and Usability. Human-Computer Interaction 19(4), 311-318 (2004)
7. Robins, D., Holmes, J.: Aesthetics and Credibility in Web Site Design. Information Processing & Management 44(1), 386-399 (2008)
8. Lynch, P.: Visual Decision Making. A List Apart: For People Who Make Websites 286 (2009)
9. Jordan, P.W.: Contemporary Trends and Product Design. CRC Press (2000)
10. Porat, T., Tractinsky, N.: The Effects of Perceived Web-Store Characteristics on Consumers' Affective States and Attitudes Towards the Store. In: Fifth Conference on Design&Emotion, pp. 27-29. (2006)
11. Mano, H.: The Structure and Intensity of Emotional Experiences: Method and Context Convergence. Multivariate Behavioral Research 26(3), 389-411 (1991)
12. Suri, G., Sheppes, G., Gross, J.J.: Predicting Affective Choice. Journal of Experimental Psychology - General 142(3). 627 (2013)
13. Suri, G., Gross, J.J.: Emotion Regulation and Successful Aging. Trends in Cognitive Sciences 16(8), 409-410 (2012)

14. Dimoka, A.: How to Conduct a Functional Magnetic Resonance (Fmri) Study in Social Science Research. MIS Quarterly 36(3), 811-840 (2012)
15. Bettman, J.R., Luce, M.F., Payne, J.W.: Constructive Consumer Choice Processes. Journal of Consumer Research 25(3), 187-217 (1998)
16. Lavie, T., Tractinsky, N.: Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. International Journal of Human-computer Studies 60(3), 269-298 (2004)
17. Tractinsky, N., Katz, A., Ikar, D.: What Is Beautiful Is Usable. Interacting with Computers 13(2), 127-145 (2000)
18. Beaudry, A., Pinsonneault, A.: The Other Side of Acceptance: Studying the Direct and Indirect Effects of Emotions on Information Technology Use. MIS Quarterly 34(4), 689-710 (2010)

Identifying Neurological Patterns Associated with Information Seeking: A Pilot fMRI Study

Javed Mostafa[1,3*], Vincent Carrasco[1,3], Chris Foster[2,3], and Kelly Giovenallo[2,3]

[1] Information Science & Biomedical Research Imaging Center, [2]Cognitive Neuroscience & Biomedical Research Imaging Center, [3]The University of North Carolina at Chapel Hill, USA
{jm,fostercm,kgio}@unc.edu,vcarrasc@live.unc.edu

Abstract. The abstract is a mandatory element that should summarize the contents of the paper and should contain at least 70 and at most 150 words. Abstract and keywords are freely available in SpringerLink.

Keywords: We would like to encourage you to list your keywords here. They should be separated by middots.

1 Introduction and Background

Searching for information online is one of the most ubiquitous activities people engage in. Search, as an independent application is highly popular and search as an embedded function appears in numerous applications and services. In fact, it can be argued that users who do not engage in online searching or perform it poorly are at a significant disadvantage in keeping pace with the everyday life demands in modern information-intensive societies. It is thus crucial that we develop a clear and evidence-based understanding of how humans conduct online searching and what constitutes their biological underpinnings.

Among various information-centric disciplines, online searching has been most intensively studied by information scientists. Research evidence gathered in the last quarter century, as illustrated in a recent comprehensive survey, was primarily influenced by behavioral models [1]. They range from basic information-gap oriented models, such as Belkin's Anomalous State of Knowledge (ASK) model [2] to more complex ones such as Pirolli's information foraging model [3]. It is only in the last four of five years that researchers started to focus more intensely on physiological and neurological evidence associated with online searching. Researchers have looked at evidence collected by eye-trackers [4] and fMRI devices [5] to establish a deeper understanding of "relevance" and how it pertains to searches and some researchers are engaged in developing and evaluating feedback systems that support iterative refinement of searching [6]. Efforts are now underway in the broader human computer interaction (HCI) community to expand the focus on neuro-physiological methods for elucidating cognitive load and stress. Contributions by information systems research-

*Corresponding author.

ers such as Riedl et al. [7] and Dimoka et al. [8] and by information scientists Maior et al. [9] and Pike et al. [10] are representative of research that attempted to link neuro-physiological approaches to HCI.

There are two primary foci in this study, namely the influence of search complexity on neuronal activations and the association between search result presentation and neuro-activation patterns. The main goal was to establish how progressively challenging search tasks influences cognition as far as it can be ascertained from brain activation patterns. Search tasks were operationalized as those typically performed on a modern search engine such as Google. Detection of neuronal activation was performed using fMRI. Studying information search with an fMRI machine is highly challenging, as issues associated with electro-magnetic interference with devices, constrained space, and movement of subjects have to be dealt with. Therefore, a secondary focus was on development and refinement of an fMRI-based research methodology for studying online searching. In the sections that follow, we will describe our pilot study method and the process for MR image acquisition and image analysis. The paper concludes with results and discussions on the main findings.

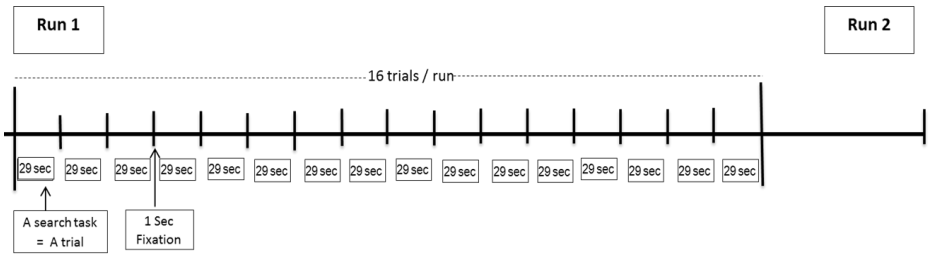

Fig. 1. An event based design for evaluating search steps conducted during fMRI sessions.

2 Method

2.1 General Design

We engaged 12 healthy subjects, age ranging between 18 and 25, without any known cognitive or neurological disorders. All were right-handed and recruited from the campus of the University of North Carolina at Chapel Hill. The basic design of the study consisted of a 2 by 2 format for the independent variables, whereby one level of the experiment focused on task type (topical vs. factual) and another level concentrated on result presentation (high precision vs. low precision). With the 2x2 design for the independent variables, the four categories of search tasks (trials) were: 1) topical and low precision, 2) topical and high precision, 3) factual and low precision, and 4) factual and high precision. It is well known that topical searching, being relatively conceptual and abstract, as compared to factual searching, is generally more challenging to perform. High precision ranking is when correct search results appear as the highest ranked items, as normally expected, and, hence, supportive of easier navigation. Low precision ranking is when correct search results are purposefully placed in

the bottom 3-4 rank, thus making the result list difficult to navigate and use. The experimental design, exploiting a mixture of task type and ranking, permitted us to investigate search difficulty on a continuum, in the following order: topical-low precision being most challenging, topical-high precision next, followed by factual-low precision, and ending with factual-high precision as the least challenging. Subjects were given an equal number of search tasks from each of the four categories (20 each, 80 total). Search performance was measured using two dependent variables: search time and search accuracy. Our hypothesis was that search performance would be systemically influenced by search difficulty, and the neural activation patterns would also be differentiated corresponding to the search performance challenges.

2.2 Tasks

Before entering the MR machine, subjects were given 5 orienting trials. A trial consisted of a screen presentation which included a search query at the top, a search box containing keywords (usually a subset of important words from the query), and a ranked list of search results. The query was either a factual or a topical type. A factual type typically required a direct and concrete answer, such as "What is the capital of Canada?" and topical search required critical reflection and response to a conceptual question such as "Who was the most influential economist in the last century?". Upon entering the MR machine, subjects were given an opportunity to ask any questions. The screen content was presented on a back-project MR-safe display and responses from subjects were collected using two MR-safe response boxes on which subjects could indicate a choice between 1-9 (inclusive) using their fingers (left pinky to right ring finger). The experimental trials and MR data were collected immediately after subjects indicated they were ready to start by pressing on a button. Each of the 80 trials lasted exactly 29 seconds with a 1 second fixation separating trials and sixteen trials were presented during each run. Trial presentation was pseudo-randomized such that no trial type was repeated more than three times in a row. A total of five, 8 minute runs were completed by each participant. Scan-time was approximately one hour. Order and content of blocks were counterbalanced across participants.

Image Acquisition: A Siemens 3 Tesla Trio imaging system. An anatomical scan was acquired for each participant using a high resolution T1 weighted MPRAGE sequence (repetition time [TR] – 1900 ms; echo time [TE] – 2.26 ms; flip angle – 9°; 192 slices, field of view [FOV] – 256; matrix - 256 x 256, 1.0 x 1.0 x 1.0 mm resolution). After the anatomical scan, 5 functional runs were acquired for each participant. A T2* weighted echo planar imaging (EPI) sequence (TR – 2000 ms; TE – 23 ms; FOV – 256 mm; flip angle – 80°) was used to collect functional images.

Analysis: Imaging data were processed using SPM 8 (Wellcome Department of Cognitive Neurology, London) run within Matlab (Matlab Mathwork, Inc., Natick, MA). For preprocessing, fMRI data were slice-time corrected for acquisition order (referenced to the first slice), then realigned and un-warped to correct for motion across

runs. Next, the images were spatially normalized by warping each participant's anatomical scan to MNI (Montreal Neurological Institute) defined standardized brain space (resampled at 2 x 2 x 2 mm), and then applying that algorithm to the EPI data. Statistical analyses were performed using the general linear model for event related designs in SPM 8. For each participant, a whole brain voxel wise analysis was conducted in which instances of a particular event type were modeled through the convolution with a canonical hemodynamic response function. Because the search task required an extended period of time to: (1) read the question, (2) read through the results, and (3) choose an appropriate response; each event was modeled from the onset of the stimulus until the participant made a response for that trial. Trials were binned according to trial type and the accuracy of the subject's response. At the fixed effects level, each condition type was contrasted with fixation. These data were then entered into a second order, random effects analysis. A 2 (type: Factual vs. Topical) x 2 (Precision: High vs. Low) x 2 (Accuracy: Correct vs. Incorrect) full factorial analysis of variance (ANOVA) was conducted using a threshold of $p < .001$ with a minimum cluster size of 5 contiguous voxels ($k \geq 5$).

3 Results and Discussion

3.1 Differentiated Activations

As anticipated, search difficulty did produce differentiated activations. A three-way ANOVA, with 2 (task type: factual vs. topical) x 2 (ranking: high precision vs. low precision) x 2 (accuracy: correct vs. incorrect), identified a significant main effect for task type, ranking, and an interaction between ranking and accuracy (Table 1). We conducted t-tests to establish the directionality of the main effects and the interaction.

3.2 Discussion

It appears that task types that are topical, particularly those of narrative nature, activate the superior temporal gyrus (STG). In our findings we noticed that the topical searches impacted the left STG and the temporal differently than the other types of searches. On the other hand factual searches generated a very different response. Because factual searches are generally easier to perform, searchers are supposed to perform them with greater confidence and assuredness. Hence, the dominance of activations observed for factual searches in the in frontal regions, known to be involved in executive function and decision making, i.e., middle frontal gyrus (MFG) and inferior frontal gyrus (IFG), and in the temporal region, associated with recognition memory and visual processing (i.e., inferior temporal gyrus and parahippocampal gyrus), fit our expectations. Factual searching depend much for directly on "acquired knowledge", instead of browsing and scrutinizing the options presented on the screen. Our observation of broad activation patterns, associated with factual searchers, ranging from bilateral MFG, bilateral inferior temporal gyrus, right parahippocampal gyrus, right IFG, to superior occipital gyrus and superior parietal lobe may partly be

representative of the participant involved in extracting a known answer rather than confirming or disconfirming an answer presented to them.

In the MR scanner the low-numbered response box (1-6) required left hand use and the high-numbered response box required the right hand (6-9). High precision correct results appear at the top, hence, they are associated with low numbers (1-5), and conversely low precision accurate results appear at the bottom, and, hence, they are associated with high numbers (6-9). The differentiated motor response may partially explain the high greater than low precision (i.e., activation in the postcentral gyrus) and low greater than high precision (i.e., activation in the precentral gyrus).

Table 1. Neural activation results from full factorial analysis of variance.

				MNI Coordinates				
Contrast	Region of Activation	Hemisphere	BA	x	y	z	t	k
Main effect of Type								
Factual > Topical	Middle Frontal Gyrus	R	6	30	18	58	6.24	2972
	Middle Frontal Gyrus	L	6	-26	4	52	5.72	4300
	Thalamus	R	N/A	8	-10	2	5.15	375
	Inferior Temporal Gyrus	L	20	-54	-52	-8	4.53	124
	Precuneus	L	7	-24	-62	40	4.48	272
	Inferior Frontal Gyrus	R	46	44	48	-10	4.24	108
	Inferior Temporal Gyrus	R	20	54	-26	-16	4.14	47
	Parahippocampal Gyrus	R	35	34	-24	-24	3.90	13
	Sub-Gryal	R	37	54	-44	-10	3.88	19
	Thalamus	L	N/A	-8	-20	12	3.80	10
	Superior Occipital Gyrus	R	19	36	-68	32	3.74	60
	Thalamus	L	N/A	-10	-12	4	3.70	9
	Superior Parietal Lobule	R	7	30	-66	48	3.70	38
	Middle Frontal Gyrus	L	10	-38	56	8	3.66	6
Topical > Factual	Superior Temporal Gyrus	L	38	-44	10	-32	5.67	291
Main effect of Precision								
High > Low	Postcentral Gyrus	R	3	38	-26	60	4.34	308
	Middle Frontal Gyrus	R	8	56	22	34	4.04	47
	Superior Frontal Gyrus	R	6	30	16	62	3.73	33
Low > High	Precentral Gyrus	L	4	-36	-24	68	6.02	882
Precision X Accuracy								
LPC > All	Postcentral Gyrus	L	2	-38	-26	68	6.41	705

Note: BA = Brodmann Area, k = cluster size, LPC = low precision correct

Similarly, the somatosensory cortex, part of the left postcentral gyrus, is also known to be involved in motor planning. The low precision correct (LPC) greater than all

other trial types and the dominant activation in the somatosensory cortex for LPC trials implies that motor planning was being carried out. Subjects in the LPC trials anticipated finding the correct results at the bottom. Hence, it is likely that they began planning to respond with their right hand while browsing and reading the top items.

This is an ongoing project. We plan to scale-up the study to include more subjects and continue to investigate the neuronal activation patterns and their associations with critical dimensions of searching.

Acknowledgement: NSF Grant #1444896 partially supported this study.

References

1. Fisher, K.E., Erdelez, S., McKechnie, L.E.F. (eds.): Theories of Information Behavior. Information Today, NY (2005)
2. Belkin, N.J., Oddy, R.N., Brooks, H.M.: ASK for Information Retrieval: Part I. Background and Theory. The Journal of Documentation 38(2), 61-71 (1982)
3. Pirolli, P.: Information Foraging Theory. Oxford University Press, NS (2009)
4. Gwizdka, J.: Characterizing Relevance with Eye-tracking Measures. In: Proceedings of the 5th Information Interaction in Context Symposium, pp. 58–67. ACM Press, NY (2014)
5. Moshfeghi, Y., Pinto, L.R., Pollick F.E., Jose, J.M.: Understanding Relevance: An fMRI Study. In: Proceedings of the European Conference on Information Retrieval, pp. 14-25. Springer, Heidelberg (2013)
6. Moshfeghi, Y., Jose, J.M.: An Effective Implicit Relevance Feedback Technique Using Affective, Physiological and Behavioural Features. In: Proceedings of the 36th international ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 133–142, ACM Press, NY (2013)
7. Riedl, R., Randolph, A., Brocke, J. vom, Léger, P.-M., Dimoka, A.: The Potential of Neuroscience for Human-Computer Interaction Research. In: SIGHCI Proceedings, paper 16. AIS (2010)
8. Dimoka, A., Pavlou, P.A., Davis, F.: NeuroIS: The Potential of Cognitive Neuroscience for Information Systems Research. Information Systems Research 22, 687-702 (2010)
9. Maior, H., Pike, M., Sharples, S., Wilson, M.L: Examining the Reliability of Using fNIRS in Realistic HCI Settings for Spatial and Verbal Tasks. In: Proceedings of CHI'15 (accepted)
10. Pike, M. F., Maior, H. A., Porcheron, M., Sharples, S. C., Wilson, M. L.: Measuring the Effect of Think Aloud Protocols on Workload using fNIRS. In: Proceedings of CHI'14 pp. 3807-3816. ACM Press, NY (2014)

Proposal for the Use of a Passive BCI to Develop a Neurophysiological Inference Model of IS Constructs

Adriane B. Randolph[1, (✉)], Élise Labonté-LeMoyne[2], Pierre-Majorique Léger[2], François Courtemanche[2], Sylvain Sénécal[2], and Marc Fredette[2]

[1] Kennesaw State University, Department of Information Systems, Atlanta, Georgia, USA
arandolph@kennesaw.edu
[2] HEC Montreal, Tech3Lab, Montreal, Quebec, Canada
{elise.labonte-lemoyne,pml,francois.courtemanche,sylvain.senecal, marc.fredette}@hec.ca

Abstract. The measurement of constructs in the field of information systems (IS) is often performed with the use of retrospective or intrusive psychometric tools that may be subject to biases. Using a passive brain-computer interface (BCI) to measure these constructs continuously in real-time without interrupting the participants would be a great addition to the toolbox of IS researchers. While the development of BCIs has been explored elsewhere, we present here a specific framework using passive BCIs to develop a neurophysiological inference model of IS constructs.

Keywords: Passive BCI · construct development · classification framework · feature extraction · cognitive load · implicit measure · NeuroIS

1 Introduction

Previous NeuroIS research has highlighted the benefits of complementing traditional explicit measures, self-declared measurements which represent the user's conscious perceptions, with implicit measures, automatic or subconscious measurements such as with electroencephalography (EEG), to enrich our understanding of information systems (IS) phenomena [1-3]. The implicit measures selected, such as EEG-based engagement indices, rely on brain signal decoding techniques developed by researchers in psychophysiology [4,5]. These brain signal decoding techniques have been incorporated into passive brain-computer interfaces (BCIs) to allow realtime feedback to a system through ongoing cognitive monitoring of the user [6]. These realtime brain signal decoding techniques may be incorporated into IS research for similar ends.

Such realtime BCI-based measures are important because they can provide information on a person's reactions to a system as they occur while he/she actually interacts with the interface. In contrast, traditional psychometric tools are used post-hoc or in aggregate form which does not provide the degree of resolution needed for truly

responsive system design. Thus, BCI-based measures can help to inform IS design as well as enable neuroadaptive systems [7].

To harness the power of passive BCIs, the IS field needs a rigorous and transparent guideline to follow. Without such guideline, researchers risk relying on a "blackbox" approach with little understanding or control of the underlying mechanisms as quite often is the case with commercial tools. More and more commercially available tools are becoming available to NeuroIS researchers but not necessarily offering more insights. For example, Emotiv, a consumer-grade EEG system, provides several built-in EEG indices but does not disclose the actual algorithm (www.emotiv.com). Other groups of neurophysiological researchers have chosen to disclose their algorithm while still protecting the methodology in patents as with work by Advanced Brain Monitoring, Inc. [8]. Building on passive BCI tools and techniques, this paper presents a framework for developing a neurophysiological inference model of an IS construct as well as a pilot study to illustrate this with cognitive load.

2 General Brain-Computer Interfaces Framework

A BCI is a system that takes human thought and translates it into a way to control a device or computer [9,10]. Based on neural control, it does not rely on any voluntary muscular input whereas most traditional interfaces require some form of voluntary muscular control for pointing-and-clicking, typing, touching, or even looking at the screen. Although BCIs have traditionally targeted users with severe motor disabilities to augment or enhance their abilities [11], researchers are increasingly looking to mainstream applications for healthy users. One such way healthy users may benefit is in the form of a passive BCI which uses brain activity "for enriching a human–machine interaction with implicit information on the actual user state" [6].

A BCI is typically thought to rely on activity generated from the central nervous system (CNS). This activity may be recorded non-invasively using a tool such as EEG from the scalp [10]. The components of a BCI include signal acquisition and processing, feature extraction and classification, a translation algorithm for converting EEG features into machine-readable format, and output of a command signal to a device as illustrated in Figure 1. For this work, the device command is not intended for active or reactive BCI use as with control purposes, but instead for passive BCI use.

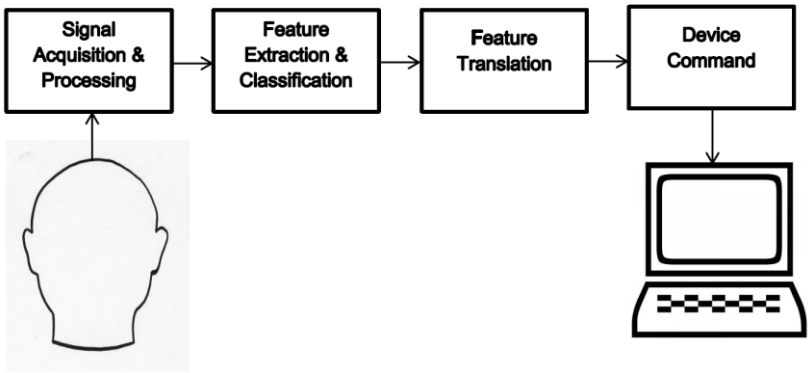

Fig 1. General BCI Framework

3 A Proposed Framework for the Development of BCI-based Measurement of IS Constructs

As Figure 1 presents the general framework for BCI studies, Figure 2 presents the specific framework used for the classification of constructs with neurophysiological measures.

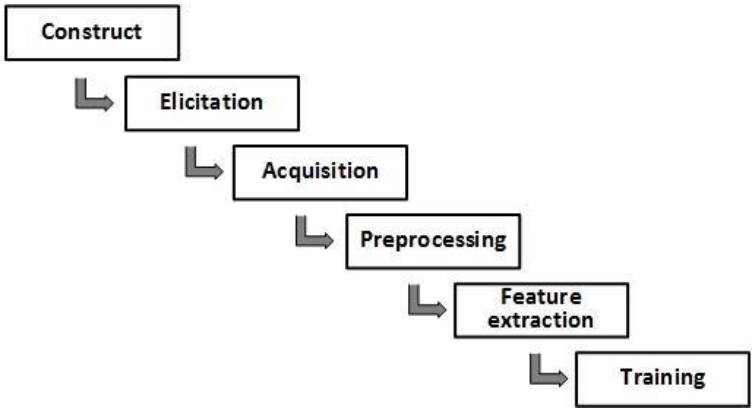

Fig. 2. Classification Framework Used with Authors' Permission [12]

Each step of this framework represents a set of actions associated with classification. At each one of these steps, the concepts of reliability, validity, sensitivity, diagnosticity, objectivity, and intrusiveness [13] should be considered.

- Construct
  - Define IS construct of interest
- Elicitation
  - Find linked neuropsychological construct with a validated elicitation task
- Acquisition
  - Data acquisition
  - Using elicitation task to put subjects in discrete levels of construct
- Preprocessing
  - Remove noise and artifacts from neurophysiological data
- Feature extraction
  - Selection of best feature corresponding to construct in neurophysiological measure [14, 15]
- Training and validation
  - Train the model
  - Three types: subject-dependency, task-dependency, lab versus in-the-wild
    - Use model to predict state in the inducing task with same individual
    - Use model to predict with other individual
    - Use model to predict in authentic IS context

4 Illustrative Study: The Development of a Cognitive Load Index for IS Research

The purpose of this illustrative study is to present the components necessary for developing a neurophysiological inference model of an IS construct and are described in the following table.

Table 1. Components of an Illustrative Experiment with the Classification Framework

Framework	Illustrative Experiment
Construct	Cognitive load • Has been previously used in IS [2] • Multiple validated tasks exist to elicit discrete levels
Elicitation	N-Back task • Strongly validated and often used to elicit cognitive load [16] • Neurophysiological tools such as EEG [16, 17] and pupillometry [18] have been used to assess cognitive load
Acquisition	50 participants in the fall of 2014 and spring of 2015
Preprocessing	Filtering, artifact removal, eye movement corrections, referencing, and frequency separation of EEG data

Framework	Illustrative Experiment
Feature extraction	• Extraction of the features from the conditioned signals. Multiple iterations will help to ensure the best feature is identified from which to construct indices • Feature conditioning to properly prepare the feature vector for the feature-translation stage
Feature translation	• Selecting a model • Parameterizing a model • Evaluating translation algorithms
Training & Validation	• Training will be completed in the summer of 2015 • A subset of the participants will be used to validate the model • An online shopping task with two levels of cognitive load requirements was also performed by the same subjects and will be used for validation in an authentic setting

5 Conclusion

Here we present a specific framework for the development of passive BCIs to measure IS constructs. The ability to monitor IS constructs in realtime with a much more detailed understanding of their evolution over time will allow for great advancements in our comprehension of the constructs themselves. In addition to cognitive load, the concept of flow is of particular interest as it has been particularly difficult to document adequately as, by definition, participants in this state lose perception of time. Other IS constructs that could be studied in this manner include satisfaction, technostress, and frustration. Once a tool is developed and validated, future work will include an exploration of the wide variety of devices and applications that could auto-adapt to these realtime cognitive inputs. For example, auto-adaptive serious games could maximize learning or shopping websites could be customized to a customer's cognitive style to reduce frustration. By unlocking the blackbox approach of realtime cognitive monitoring of users, researchers may gain greater insights into IS design.

References

1. Ortiz de Guinea, A., Titah, R., Léger, P.-M.: Measure for Measure: A Two Study Multi-trait Multi-method Investigation of Construct Validity in IS Research. Computers in Human Behavior 29(3), 833-844 (2013)

2. Ortiz de Guinea, A., Titah, R., R., Léger, P.-M.: Explicit and Implicit Antecedents of Users' Behavioral Beliefs in Information Systems: A Neuropsychological Investigation. Journal of Management Information Systems 30(4), 179-210 (2014)
3. Tams, S., Hill, K., Ortiz de Guinea, Ana, Thatcher, J., Grover, V.: NeuroIS - Alternative or Complement to Existing Methods? Illustrating the Holistic Effects of Neuroscience and Self-Reported Data in the Context of Technostress Research. Journal of the Association for Information Systems 15, 723–753 (2014)
4. Johnson, R.R., Popovic, D.P., Olmstead, R.E., Stikic, M., Levendowski, D.J., Berka, C.: Drowsiness/alertness Algorithm Development and Validation Using Synchronized EEG and Cognitive Performance to Individualize a Generalized Model. Biological Psychology 87(2), 241-250 (2011)
5. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic System Evaluates Indices of Operator Engagement in Automated Task. Biological Psychology 40(1), 187-195 (1995)
6. Zander, T.O., Kothe, C.: Towards Passive Brain-computer Interfaces: Applying Brain-computer Interface Technology to Human-machine Systems in General. Journal of Neural Engineering 8(2), Article 025005 (2011)
7. Vom Brocke, J., Riedl, R., Léger, P.-M.: Application Strategies for Neuroscience in Information Systems Design Science Research. Journal of Computer Information Systems 53(3), 1-13 (2013)
8. Berka, C., Levendowski, D.J., Westbrook, P.R., Davis, G., Johnson, R., Popovic, D.: Interactive Psychophysiological Profiler Method and System, U.S. Patent No. 20100292545. Nov 18 (2010)
9. Nijholt, A., Tan, D.: Brain-computer Interfacing for Intelligent Systems. Intelligent Systems, IEEE 23(3), 72-79 (2003)
10. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-Computer Interfaces for Communication and Control. Clinical Neurophysiology 113(6), 767-791 (2002)
11. Pfurtscheller, G., Müller-Putz, G.R., Scherer, R., Neuper, C.: Rehabilitation with Brain-Computer Interface Systems. IEEE Computer, 58 - 65 (2008)
12. Courtemanche, F., Dufresne, A., LeMoyne, E.L.: Multiresolution Feature Extraction During Psychophysiological Inference: Addressing Signals Asynchronicity. Physiological Computing Systems - LNCS 8908, 43-56 (2014)
13. Riedl, R., Davis, F.D., Hevner, A.R.: Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement. Journal of the Association for Information Systems, 15(10), Article 4 (2014)
14. Guyon, I., Elisseeff, A.: An Introduction to Feature Extraction. In: Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A. (eds.) Feature Extraction: Foundations and Applications, pp. 1-25. Springer, Berlin, Heidelberg (2006)
15. Guyon, I., Elisseeff, A.: An Introduction to Variable and Feature Selection. The Journal of Machine Learning Research 3, 1157-1182 (2003)
16. Gevins, A., Smith, M.E.: Neurophysiological Measures of Cognitive Workload during Human-computer Interaction. Theoretical Issues in Ergonomics Science 4(1-2), 113-131 (2003)
17. Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P., Rao, R.P.N.: Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph. In: Proceedings of

the SIGCHI Conference on Human Factors in Computing Systems, pp. 835-844. ACM (2008)

18. Buettner, R., Schauer, S., Maier, C., Eckhardt, A.: Towards Ex Ante Prediction of User Performance: A Novel NeuroIS Methodology Based on Real-Time Measurement of Mental Effort. In: Proceedings of Hawaii International Conference on System Sciences 2015, pp. 535-542. AIS (2015)

Emotion is not what you think it is: Startle Reflex Modulation (SRM) as a measure of affective processing in NeuroIS

Peter Walla[1,2 *], and Monika Koller[3]

[1] The CanBeLab, Department of Psychology, Webster Vienna Private University, Praterstrasse 23, 1020 Vienna, Austria
peter.walla@webster.ac.at
[2] Functional Neuroimaging Lab, School of Psychology, Faculty for Science and Information Technology, University of Newcastle, Newcastle, NSW, Australia
peter.walla@newcastle.edu.au
[3] Department of Marketing, Vienna University of Economics and Business, Vienna, Austria
monika.koller@wu.ac.at

Abstract. Emotion is a widely used term in various different fields. The problem is that across and even within those fields scholars are not sharing a common understanding of it. This strongly counterproductive situation hinders ongoing progress and might even lead to false understandings.
This conceptual paper offers a solution and also introduces a method called startle reflex modulation (SRM). It has been described since the late 80s in the human literature and is widely used in emotion research to measure raw affective responses. Meanwhile, besides in the frame of basic science studies it has also been applied to clinical and most recently even industry-relevant topics.
It is suggested that SRM does indeed represent a highly valuable new approach to quantify affective processing in the context of NeuroIS (e.g. technology acceptance). Often, self-reported affect differs from objectively measured affect.

Keywords: affective processing · emotion · startle reflex modulation · non-conscious · attitude · NeuroIS

1 Introduction

In 1981, two authors [1] reported in their paper that they found no less than 92 different definitions of "emotion" in a large variety of literature sources. They put all definitions into a package of 11 categories and by taking a closer look at all those definitions it seems as if the human brain is doing nothing but emotion. This is of course not true and leads to wide frustration, while one would certainly think that in 2015 we should know better. In 2013, two other authors [2] put it in a more recent perspective: "*If one is after a definition of emotion it turns out that almost as many definitions*

exist as textbooks are available. Just as some scholars would call the engine of a car "car", while others would call driving a hunk of sculpted metal a "car" without realizing that the engine is essential for driving, but it actually is a separate thing deserving its own label".

What is the solution to this problem? The solution is to take recent scientific outcome and new theories into account and then simply use different labels for different aspects of emotion-related processes and/or behaviors. Referring to the above mentioned car-analogy, only if we label things properly the mechanic knows that one is talking about the wheels of a car, someone else about the engine, while yet another one talks about driving. For the purpose of this paper, affective processing is understood as neural activity of subcortical structures that reflects valence-aspects of a stimulus ("how" is a stimulus). It is purely non-conscious and feeds into any decision making process. It is strongly linked with motivation and thus forms a basic driver for us to get active. Affective processing can cause further bodily responses that can lead to conscious feelings (e.g. the feeling of fear). Finally, one or more emotions can be generated that are meant to be expressions, which means they are a consequence of muscle contractions. After all, the latin word "emovere", which forms the word basis for "emotion" means to "take out" or in other words to express. Emotion is not neural activity and thus it is not information processing, nor is it anything felt, it simply is the behavioral output of information processing, while information in this case is affective. Hence, Walla & Panksepp [2] gave their chapter the title "*Neuroimaging helps to clarify brain affective processing without necessarily clarifying emotions*".

We are aware of the difficulty to change a long tradition and history, but we believe that simple solutions are better solutions.

In the frame of this paper we focus on affective processing and an effort is made to introduce an unfortunately neglected method to the field of NeuroIs. It is called startle reflex modulation (SRM) and it is better than any expensive and sophisticated brain imaging technology when it comes to quantify affective processing, which in other words means to measure grade of pleasantness. That is exactly what most people are after. Traditionally, explicit rating performance is used to infer pleasantness, but crucially we here show a few exemplary cases that highlight discrepant findings when self-reported responses are compared with objective responses. The brain knows more than it admits, especially with respect to affective information.

2 Affective Information Processing - An Evolutionary Perspective

Information processing has cognitive and affective aspects and from an evolutionary perspective the more important driving force for behavior adaptation is affective processing. In terms of pure survival it is obviously more important to have an idea about whether you can approach or you should rather avoid something out there in comparison to knowing what it is. At an early stage during our own lives we don't even have any sophisticated cognitive capacity, but evolution provided us already with the capacity to affectively evaluate our environment – we are capable of affective infor-

mation processing. This type of information processing evolved first to detect potential harm and to find appetitive sources in an ever-changing environment. Only later, evolution equipped organisms with a cognitive system that finally allowed language to come into existence. Since affective processing evolved earlier than any cognitive development, it did obviously not depend on language. In other words, affective information was never meant to be verbalized. Only semantic information, the very basis for cognition, is designed to be put into words. As human beings, we all have the cognitive ability to use words to supposedly verbalize even affective information. However, due to the non-cognitive and also non-conscious nature of it, words may terribly fail to do so. And even worse, words can easily be used to intentionally misinform others to leave a good impression or to help them with their survey investigation.

3 Affective Information Processing – A Measurement Perspective

In search for the most reliable method to quantify affective responses as in how pleasant or unpleasant something is we came to the conclusion that even by taking all brain imaging technology into account it actually is startle reflex modulation (SRM).

After pioneering investigations using rodents, it was found that humans too demonstrate a modulated startle reflex as a reaction to affective state [3,4]. Since then, the magnitude of an eye-blink response to loud and short acoustic white noise (the startle probe) has been taken as a measure of raw affect [5,6,7,8]. Crucially, it has been shown that even rapid changes in affective content (valence) related to so-called lead stimuli are reflected in eye-blink amplitudes as startle responses, which thus results in modulated responses, hence the name startle reflex modulation. In summary, the more positive the state of affect (while being startled) the more reduced the startle response and vice versa. Besides the objective nature of this method, its main advantage is its independence from cognitive information processing. The fact that startle responses adapt quickly to changing valence, together with their independence from cognitive influence, makes them an ideal tool with which to quantify pure affective processing on a fine-graded scale and to potentially reveal discrepancies between self-reported and objectively measured affective processing [9,10,11].

4 Discrepancies between self-reported and objective data

Having underlined the relevance of being able to validly measure affective information processing in the theoretical framework section above, we will in the following introduce a few selected examples of how startle reflex modulation as a measure of deep inner affect can reveal what does not feed into explicit responses.

In IS virtual realities have long been a hot topic [12]. Recently, first attempts were made to tap into their non-conscious aspects. Pre-defined walking tracks have been used via Google StreetView as a virtual setting while measuring affective processing associated with urban environments [13]. Participants had to navigate through six

different districts in Paris that were chosen on the basis of varying median real estate price while their startle responses were recorded. After the recording session participants were asked to rate the subjective pleasantness of each district. For the most expensive and the cheapest district explicit ratings matched up with startle responses. However, for the rest this correlation was not evident. In fact, the second most expensive district elicited a surprisingly enhanced startle response, which indicates negative affect while virtually walking through it. Such discrepancies between subjective and objective measures of affective processing might become a critical field of application in IS and NeuroIS [6].

A further example is about product aesthetics. Three bottles that differed only in shape were presented to males and females while asking them to rate them with respect to attractiveness while simultaneously their startle responses were recorded. It has been shown that one specific bottle (out of the three) elicited significantly enhanced startle responses in males compared to females, although both females and males rated this particular bottle as equally least attractive among all [14]. This is another incidence of a discrepancy between self-reported and objectively measured affect.

A study about self- versus non-self-referenced affective content revealed that participants rated pleasant images as more pleasant and unpleasant images as more unpleasant when these were self-referenced [15]. Respective startle responses though were potentiated related to pleasant self-referenced images indicating greater affective negativity, which is in contrast to greater self-reported positivity. Only for the unpleasant valence category did the authors find a match between self-report and startle responses as objective measures.

Finally, virtually driving through a dark tunnel does not always match simultaneously registered startle responses [16] and so do increasingly angrier faces not necessarily lead to similar linear changes in explicit responses versus startle responses [17].

5 Conclusion and Discussion

The main focus of this paper was to conceptually reflect on affective information processing as the underlying basis of emotions that in turn might be best understood as behavioral output. Based on an evolutionary and biological discussion of these concepts, startle reflex modulation represents a reliable approach to measure affective information processing as in grade of pleasantness. In NeuroIs, for example technology acceptance is a big field [18] and startle reflex modulation could now significantly contribute to it by helping to get access to deep inner affect that finally forms the basis for any following decision making process.

References

1. Kleinginna, P.R.Jr., Kleinginna, A.M.:A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition. Motivation and Emotion 5(4), 345-379 (1981)
2. Walla, P., Panksepp, J.: Neuroimaging Helps to Clarify Brain Affective Processing without Necessarily Clarifying Emotions. In: Fountas, K.N. (ed.) Novel Frontiers of Advanced Neuroimaging, pp. 93-118. INTECH (2013)

3. Bradley, M. M., Lang, P. J., Cuthbert, B. N.: Emotion, Novelty, and Startle Reflex: Habituation in Humans. Behavioural Neuroscience 107(6), 970-980 (1993)
4. Dawson, M.E., Schell, A.M., Böhmelt, A.H.: Startle Modification: Introduction and Overview. In: Dawson, M.E., Schell, A.M., Böhmelt, A.H. (eds.) Startle Modification. Implications for Neuroscience, Cognitive Science, and Clinical Science, pp. 6-18. Cambridge University Press, Cambridge (1999)
5. Walla, P., Brenner, G., Koller, M.: Objective Measures of Emotion Related to Brand Attitude: A New Way to Quantify Emotion-related Aspects Relevant to ;arketing. PloS ONE 6(11), Article ID e26782 (2011)
6. Koller, M., Walla, P.: Measuring Affective Information Processing in Information Systems and Consumer Research – Introducing Startle Reflex Modulation. In: ICIS 2012 Proceedings. AIS (2012)
7. Lyons, G.S., Walla, P., Arthur-Kelly, M.: Toward Improved Ways of Knowing Children with Profound Multiple Disabilities (PMD): Introducing Startle Reflex Modulation. Developmental Neurorehabilitation 16(5), 340-344 (2013)
8. Walla, P**.,** Koller, M., Meier, J.: Consumer Neuroscience to Inform Consumers - Physiological Methods to Identify Attitude Formation Related to Over-consumption and Environmental Damage. Frontiers in Human Neuroscience, 20 May (2014)
9. Walla, P., Richter, M., Färber, S., Leodolter, U., Bauer, H.: Food Evoked Changes in Jumans: Startle Response Modulation and Event-related Potentials (ERPs). Journal of Psychophysiology 24, 25-32 (2010)
10. Walla, P., Rosser, L. Scharfenberger, J. Duregger, C., and Bosshard, S.: Emotion Ownership: Different Effects on Explicit Ratings and Implicit Responses. Psychology 3A, 213-216 (2013)
11. Mavratzakis, A., Molloy, E., Walla, P.: Modulation of the Startle Reflex during Brief and Sustained Exposure to Emotional Pictures. Psychology 4, 389-395 (2013)
12. Spiers, M.V., Sakamoto, M., Elliott, R.J., Baumann, S.: Sex Differences in Spatial Object-Location Memory in a Virtual Grocery Store. CyberPsychology & Behavior 11(4), 471-473 (2008)
13. Geiser, M., Walla, P.: Objective Measures of Emotion during Virtual Walks through Urban Neighbour-hoods. Applied Sciences 1, 1-11 (2011)
14. Grahl A., Greiner, U., Walla, P.: Bottle Shape Elicits Gender-specific Emotion: A Startle Reflex Modulation Study. Psychology 7, 548-554 (2012)
15. Walla, P**.,** Rosser, L. Scharfenberger, J. Duregger, C., Bosshard, S.: Emotion Ownership: Different Effects on Explicit Ratings and Implicit Responses. Psychology 3A, 213-216 (2013)
16. Mühlberger, A., Wieser, M.J., Pauli, P.: Darkness-enhanced Startle Responses in Ecologically Valid Environments: A Virtual Tunnel Driving Experiment. Biological Psychology 77, 47-52 (2008)
17. Dunning, J.P., Auriemmo, A., Castille, C., Hajcak, G.: In the Face of Anger: Startle Modulation to Graded Facial Expressions. Psychophysiology 47, 874-878 (2010)
18. Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13(3), 319–340 (1989)

Measuring flow using psychophysiological data in a multiplayer gaming context

Marie-Christine Bastarache-Roberge, Pierre-Majorique Léger, François Courtemanche, Sylvain Sénécal and Marc Fredette

HEC Montreal, Montreal, Canada
{marie-christine.bastarache-roberge, pierre-majorique.leger, francois.courtemanche, sylvain.senecal,marc.fredette}@hec.ca

Abstract. Flow is a desirable state where an individual is focused and satisfied. Traditional flow models are based on an individual's skills and the challenges he faces. The objective of this ongoing research is to investigate, in a gaming context, how a player's and his teammate's personality and neurophysiological reactions can contribute in explaining a player's flow assessment. Our preliminary results show that adding these measures significantly increases the performance of predicting flow models.

Keywords: Flow in team · Games · Multiplayer · Neurophysiological data

1 Introduction

Flow is a mental state in which a person is fully focused on, involved in, and enjoying the task at hand. The flow state is defined as an optimal experience in which "the person feels simultaneously cognitively efficient, motivated, and happy" [1] (p.277). In this state of focused immersion, heightened enjoyment, and temporal dissociation, players feel intrinsically motivated and in total control of the game environment [2].

Although prior research has extensively investigated individual flow, research on flow in teams is nascent [3]. Recently, researchers have started to investigate flow episodes involving more than one person [4,5]. To date, flow experiences in teams have been mostly studied using psychometric tools in sports and other social contexts such as music and dance, or even in the context of highly engaging conversations [6].

Prior research found that using participants' personality or physiological data can improve the prediction of flow. According to Hanin [7], both psychological and physiological dimensions are essential to understand cognitive and emotional conditions underlying a flow state. Thus, Peifer [8] proposes an integrative definition of flow experience which states that flow is a positive valence state (affective component), resulting from an activity that has been appraised as an optimal challenge (cognitive component), characterized by optimized physiological activation (physiological component).

Recent research in NeuroIS have proposed to enrich current flow measures by capturing automatic and psychophysiological measures in conjunction with self-reported

measures of flow of an individual [9]. This paper proposes to extend this model to a group environment. Building on the theory of emotional contagion [10], this research develops a model for inferring individual flow from psychophysiological states and personality traits of participants involved in the task.

We tested our model in a laboratory experiment conducted with 88 subjects who played a total of 120 games in teams of two. Preliminary results show that the participant's flow and his teammate's flow are strongly correlated. Furthermore, the accuracy of our individual flow inference model increases when the physiological and personality data of both players are included in the model.

2 Method

2.1 Participants

Eighty-eight participants (31 females, 57 males) were recruited and took part in the experiment over a period of two weeks. Each experiment lasted 2 hours and participants received a compensation of $25. All participants were university students.

2.2 Apparatus

Team Fortress 2 (Valve Corporation, Bellevue), a first person shooter game, was used for the experiment. This game was selected because of the possibility to manipulate the game's difficulty levels and extract players' actions (e.g., successful or missed shots, games won) from the game logs with the aid of an additional plug-in (Supstats2 by F2). The HEXACO personality test assessing Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience personality factors [10] was completed by participants before the experiment. A self-reported flow measurement scale [11] was filled out after each game, along with a final questionnaire with other measures: gaming skill, age and sex. During the experiment, the following psychophysiological data was recorded: cardiac activity, respiration, electrodermal activity with a Biopac MP-150 amplifier (Biopac, Goleta) and emotional valence and arousal with the Facereader system (Noldus, Wageningen). Following, neurophysiological data was artefacted and synchronized with the players' physiological data and self-reported flow.

2.3 Procedure

Subjects were invited to the lab in groups of 3 or 4 and alternated between playing Team Fortress 2 and watching games of the other players on their screen (See Figure 1). Players were all in the same room, but could not directly see each other. They could communicate using their headset, which simulated an online environment. The first game was a practice game, and all participants played in cooperation against the computer. Afterward, only 2 players played per game, for a total of 6 real games per experimental session. After each game played, subjects completed a flow question-

naire. Teams were randomly assigned to various difficulty level combinations (easy, normal, hard). At the end of the experiment, subjects filled out the final questionnaire. In total, about 120 games were played.

Fig. 1. Experimental setting[1]

3 Results and concluding comments

In order to measure the influence of independent variables on self-reported individual flow, a series of multivariate linear regressions were performed. Each bloc of measures contained a certain number of variables that were then selected using a stepwise procedure. The first regression model was based on traditional flow measures and used only skill and difficulty as independent variables. It explained 25,55% of the flow variance (Table 1). When the player's personality and neurophysiological measures were added as independent variables (Model 2), the variance explained increased to 63,58%. Finally, when the partner's skill, personality and neurophysiological measures were added as independent variables in the regression model (Model 3), the variance explained increased to 68,08%. We can also see in Table 1 that each addition of variables also improved the R^2 adjusted for the number of variables in the model.

In addition, we estimated the models' ability to predict the flow of a player (1 to 7 scale) using a 10-fold crossed-validation algorithm. We can see in Table 1 that the addition of new variables in models 2 and 3 lead to an improved prediction accuracy of the flow.

[1] Note that the picture doesn't show the physiological equipment and the black blinds separating the participants.

Table 1. Regression results explaining individual flow

Models	Independent variables	Est. pr. error	R2	Adj R^2
1	Skill Difficulty	0,761	0.26	0.25
2	Skill Difficulty Personality Neurophysiological	0,431	0.64	0.61
3	Skill Difficulty Personality Neurophysiological Teammate personality Teammate neurophysiological Teammate skill	0,375	0.68	0.66

Based on these encouraging results, we are in the process of further analyzing the data to identify a parsimonious set of flow antecedents that maximize the explained flow variance. Moreover, additional statistical validation is necessary to explore and address, if needed, multicollinearity issues.

Reaching flow states while playing games is an important issue for game designers, thus gaining a better understanding of the variables contributing to flow is of the utmost importance. Our preliminary results show that a player's psychophysiological state during the game has an influence on his flow perception. This real time information, for instance captured through sensors on game controllers, could help games adjust their difficulty level in order to maximize flow. Results also show that the teammate's personality and psychophysiological state influence a player's flow assessment. This leads to interesting implications. For instance, instead of only being based on skills, player pairing could also be based on players' personality in order to increase the probability of flow episodes in multiplayer games where remote player pairing is needed. Results of this research could eventually be tested in serious gaming groups in order to maximize flow experiences of learners in group contexts.

In conclusion, this research contributes to flow research by improving our understanding of individual flow antecedents in the group contexts, which is a nascent research stream in flow research. To our knowledge, this research is the first to investigate how a teammate's personality and psychophysiological states influence a person's flow in a multiplayer gaming context.

References

1. Moneta, G.B., Csikszentmihalyi, M.: Models of Concentration in Natural Environments: A Comparative Approach Based on Strems of Experimential Data. Social Behavior and Personality: an International Journal 27(6), 603-637 (1999)
2. Agarwal, R., Karahanna, E.: Time Flies when you're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. MIS Quarterly 24(4), 665-694 (2000)
3. Walker, C.J.: Experiencing Flow: Is Doing it Together Better than Doing it Alone? The Journal of Positive Psychology 5(1), 3-11 (2010)
4. Heyne, K., Pavlas, D., Salas, E.: An Investigation on the Effects of Flow State on Team Process and Outcomes. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting 55(1), pp. 475-479. SAGE Journals (2011)
5. Léger, P.M., Sénécal, S., Aubé, C., Cameron, A.F.: The Influence of Group Flow on Group Performance: A Research Program. In: Proceedings of the Gmunden Retreat on NeuroIS 2013, p. 13 (2013)
6. Csikszentmihalyi, M., Rich, G.: Musical Improvisation: A System Approach. In: Sawyer, R.K. (ed.) Creativity in Performance pp. 43-66. Ablex: Greenwich, Connecticut (1997)
7. Hanin, Y.: Emotions in Sport. Human Kinetics, Champaign, Illinois (2000)
8. Peifer, C.: Psychophysiological Correlates of Flow-Experience. In: Engeser, S. (ed.) Advances in Flow Research, pp. 139-164. Springer, New York (2012)
9. Léger, P.M., Davis, F.D., Cronan, T.P., Perret, J.: Neurophysiological Correlates of Cognitive Absorption in an Enactive Training Context. Computers in Human Behavior 34, 273-283 (2014)
10. Barsade, S.G.: The Ripple Effect: Emotional Contagion and its Influence on Group Behavior. Administrative Science Quarterly 47(4), 644-675 (2002)
11. Lee, K., Ashton, M.C.: Psychometric Properties of the HEXACO Personality Inventory. Multivariate Behavioral Research 39, 329-358 (2004)
12. Engeser, S., Rheinberg, F.: Flow, Performance and Moderators of Challenge-skill Balance. Motivation and Emotion 32(3), 158-172 (2008)

Using a Cognitive Analysis Grid to Inform Information Systems Design

Laurence Dumont[12,], Gabrielle Chénier-Leduc[12], Élaine deGuise[2], Ana Ortiz de Guinea[1], Sylvain Sénécal[1], Pierre-Majorique Léger[1]

[1] Tech3Lab, HEC Montréal, Montréal, Canada
{laurence.dumont,gabrielle.chenier-leduc,ana.ortiz-de-guinea,sylvain.senecal,pml}@hec.ca
[2] Psychology Department, Université de Montréal, Montréal, Canada
elaine.de.guise@umontreal.com

Abstract. Following our first conceptualization of a cognitive analysis grid (CA grid) for IS research in 2014, the CA grid was improved and tested in a proof of concept manner. The theory and application of this method are briefly explained, along with lessons learned from a first experiment. The next steps in the validation of this method include applying it to a wider group of naïve participants. This will allow to draw statistical parallels between the cognitive demand of the interface and the performance of the users based on their cognitive profile. Ultimately, this technique should be useful both in NeuroIS research and user experience (UX) tests to guide hypotheses and explain user's performance.

Keywords: Cognitive Psychology · UX · Cognitive Demand · Pupil · Workload.

1 Background

Understanding interactions between interfaces and users asks for a theoretically valid conception of the user cognitive and emotional state [1]. The use of behavioral and neuroscientific methods by the NeuroIS community shows advancement in this direction and have brought forward findings that would have been impossible before [2]. However, there is still a need to build a systematic bridge between cognitive psychology and use of IS. The importance of individual characteristics is highlighted in the task-technology fit (TTF) model [3], but most attempts at operationalizing this part of the model have focused on constructs such as personality [4], computer self-efficacy [5] or various perceptual components of the technology acceptance model. The IS literature generally assumes that the cognitive abilities of users are all equal. Most research relies on the perception of cognitive processes [6] but to our knowledge, there is yet to have an integration of neuropsychological evaluation of cognitive functions and the use of IS.

Last year [7], we conceptualized how this bridge could be built. By dividing the IS interaction in steps going from the presentation of input from the interface to the output provided by the user, and by attributing cognitive processes to each steps, a Cognitive Analysis grid (CA grid) was proposed.

This CA grid gives us a profile of which cognitive functions are most solicited during a user interaction with a system. This and the user's cognitive profile can then be integrated to interpret the user's behavior. The objective of this paper is to provide a first pilot study of the CA grid with two subjects in a task that involves multiple steps. At terms, the CA grid could be a useful tool for NeuroIS researchers and user experience (UX) designers, to adapt a task to individual capacities and increase accessibility of IS to populations with special needs.

2. The CA Grid: A Revised Version

Table 1. Components, Categories and Content of the CA grid

Component	Category	Content		
Reception	Mode	Visual	Auditory	Other
	Content	Verbal	Non-Verbal	
	Response Orientation	Relevant Info.	Actionable Content	
		Non-Relevant Info.	Feedback	
Thinking		Working Memory Inhibition Set Shifting Fluency	Planning Insight Social Cognition	
Response	Mode	Motor, Vocal or Other		
	Content	Verbal or Non-Verbal		

Three main differences emerge when comparing the first version of the grid [7] and the content of the present grid in Table 1.

First, we removed the "Memory" component since it was impossible to judge to which memory system each step was related in a way that is systematic for every user[1].

Second, we added the "Response Orientation" category in the "Reception" since we had difficulty to accurately qualify the affordance relative to the content of the task, which is closely related to the cognitive demand.

Third, the theory we used to subdivide the "Thinking" component has been changed so it could be directly linked to measures of cognitive functions. The new theory, summed up in the "Thinking" component of Table 1, is based on executive functions, which originated from Baddeley & Hitch's original theory of cognition [8].

[1] Memory remains an important component of both the task and the cognitive profile of the individual, but since the content of every user's memory is different; it is difficult to make valid attributions at the level of the task.

In this theory, different cognitive modules are controlled by a central executive. This central executive has been divided in executive functions in further theoretical models. The most accepted model [9] separates three main functions: alternating between sets of rules (Set Shifting), updating working memory content (Working Memory) and inhibition of dominant responses (Inhibition). These three functions along with four other commonly evaluated functions – fluency, insight, planning, and social cognition – can be assessed using the NIH-EXAMINER, a 60 minute battery of cognitive tests [10] which provides an individual cognitive profile of the user that we can use to compare with the thinking component of the CA grid.

3. Testing the CA grid: A pilot study

3.1 Method

Participants. In order to conduct a preliminary test of the implementation of the CA grid in a user test, we asked two research assistants of the Tech3Lab to participate in a pilot experiment and to give elaborated feedback about their experience. They were both 24 year old males with similar education levels. These assistants were not directly involved in the CA grid development, but have been acquainted with the underpinnings of the project.

Material. A Tobii X60 (Sweden) eyetracker monitored their gaze and pupil dilatation. Participant's cognitive workload in the task was evaluated using average pupil dilatation while undergoing each task, a large pupil being associated with high cognitive load [11,12].

Experimental Protocol. On a first session, participants completed the cognitive tests of the French version of the NIH-EXAMINER [13]. This allowed to obtain a normative evaluation of their cognitive abilities on each function.

The task we chose was an online grocery task, which allowed sufficient different components in the task and seemed to involve varied cognitive functions. The participants had to buy the content of a 7-item list for 4 people that were coming over for dinner. They had 30 minutes to complete the process until they had to enter a credit card number. They could ask the experimenter if they had forgotten something on the list, so that their working memory would not be overloaded by always trying to remember what they had to buy. Furthermore, the main author gathered their feedback on the experiment.

3.2 Protocol to Use the CA grid

1. Evaluating the cognitive demand of the interface. The cognitive demand of the navigation in the website has to be assessed beforehand by experts in order to determine how each cognitive function should be solicited during the interaction.

To evaluate the cognitive demand of the navigation in the website, it must be divided in tasks having each a concrete goal. Each of these tasks has to be further divided in steps that start with the reception of stimuli and end with a response from

the user. In the online grocery, the two first tasks identified were “Accessing the grocery store” and “Signing up”. The steps to “Access the grocery store” were “Identifying the access to the store” and “Decide if you sign up or shop” (which the interface prompts the user to do). Table 2 shows this division between tasks and steps.

Each of these steps is then coded according to the components shown in Table 1. The Reception component is coded based on content analysis of the interface. In the example stated above, the modality would be completely visual and the content (verbal or non-verbal) would be assessed using the proportion of text to image. Response orientation of each block of content can be assessed and averaged for each task. For example, the “Accessing the grocery store” had a score of 16% relevant information, 72% non-relevant information, 7% actionable content and 5% feedback.

Table 2. Example of a completed grid, with workload and cognitive profiles

	Working Memory	Inhibition	Set Shifting	Fluency	Planning	Insight	Social Cognition	Workload[1]	
Access the Grocery Store								0,69	0,30
Identify the access to the store									
Decide if you sign up or shop									
Signing Up								0,39	0,53
Fill up the form									
Submit the form									
Choosing the delivering store									
Completion feedback									
Shopping Cart								-0,21	-0,04
Reviewing the content of the cart									
Changing the quantity of an item									
Deleting an item from the cart									
Reviewing the subtotal									
Place the order									
Search Bar									1,61
Writing a query									
Delivery								1,18	0,42
Change store									
Reserve a moment for delivery									
Cognitive Profile[2]									
Participant 1	59%	63%	87%	29%	43%	90%	95%		
Participant 2	65%	44%	84%	78%	50%	70%	92%		

[1.] The light gray represents average workload and the darker gray represents high workload
[2.] Lighter gray represents high capacities, medium gray average capacities and dark gray under average capacities

The coding of the Thinking component requires two experts with significant knowledge in cognitive psychology in order to ensure reliability. They have to identify from one to three dominant functions in each step. In our pilot, the two first authors assessed the components of the Thinking component together and validated

their judgement and discrepancies with the third author. This coding effort results in a portrait of the cognitive functions involved in the task, as shown in Table 2.

The Response component analyses the source of the user's relevant responses at every step. All of the responses were motor (either mouse clicks or typing). Mouse clicks were considered non-verbal and typing was considered verbal.

Reception and response components were not included in Table 2 given clarity and space issues. Only five of the eight tasks are shown for the same reasons.

Evaluating the users' baseline cognitive capacities. In the process of assessing TTF, the objective skills of the user should also have an effect on the interaction. For instance, a task that asks a lot of planning from the user will not be as demanding for someone who has outstanding planning skills compared to an average person. The NIH-EXAMINER allows the assessment the user's cognitive profile and could help account for inter-individual differences in using interfaces.

Our participants' cognitive profiles can be found at the bottom of Table 2 and important differences in Fluency and Insight can be highlighted.

Evaluating the users' cognitive performance in the task. Once the users are interacting with the interface, the amount of cognitive resources involved – hence the workload – will vary depending on the previously mentioned variables. This can be assessed using different tools including behavioral analysis, pupil dilatation, and different electroencephalogram frequency ranges [14,15].

In our pilot evaluation, we used pupil dilatation as a measure of cognitive load. Data was first standardized using Z scores for each participant in order to allow comparisons. The online shopping task was then divided in the eight main tasks and an average pupil dilatation measure was obtained for each of these tasks. These results are shown in the right portion of Figure 1 and are classified as high (0.5 SD above average), medium (+/- 0,5SD around average) or low (0.5SD under average). Discrepancies between participants can be found in five of the eight tasks.

4. Future Directions

In future studies, larger sample sizes will allow statistical connexions between discrepancies in cognitive profiles of the users (as measured by the NIH-EXAMINER), different task cognitive demand (as measured by the CA grid), and workload (as measured either by pupil dilatation or EEG).

It will also be possible to start to assess the reception and response skills of the users. This will be particularly useful in adapting IS to populations with special needs such as hearing or visual impairments, or motor difficulties.

Also, we are developing a method to combine the tasks' demand and the users' cognitive profile to create models of each user's performance. These models will be compared to experimental evidence of the cognitive performance. A fit between a model of performance and actual performance would validate the CA grid method.

Ultimately, validating the CA grid will provide NeuroIS researchers and UX designers with a useful tool. It will allow rigorous assessment of task to person fit,

help explain individual discrepancies in user's behaviour and increase theoretical integration between NeuroIS and cognitive psychology.

References

1. Ortiz de Guinea, A., Titah, R., Léger, P.-M.: Explicit and Implicit Antecedents of Users' Behavioral Beliefs in Information Systems: A Neuropsychological Investigation. Journal of Management Information Systems 30(4), 179-210 (2014)
2. Loos, P., Riedl, R., Müller-Putz, G.R., vom Brocke, J., Davis, F.D., Banker, R.D., Léger, P.M.: NeuroIS: Neuroscientific Approaches in the Investigation and Development of Information Systems. Business & Information Systems Engineering 2(6), 395-401 (2010)
3. Goodhue, D.L., Thompson, R.L.: Task-technology Fit and Individual Performance. MIS Quarterly 19(2), 213-236 (1995).
4. Gu, L., Wang, J.: A Study of Exploring the "Big Five" and Task Technology Fit in Web-based Decision Support Systems. Issues in Information Systems 10(2), 210-217 (2009)
5. Strong, D.M., Dishaw, M.T., Bandy, D.B.: Extending Task Technology Fit with Computer Self-efficacy. ACM SIGMIS Database 37(2-3), 96-107 (2006)
6. Ortiz de Guinea, A., Webster, J.: An Investigation of Information Systems Use Patterns: Technological Events as Triggers, the Effect of Time, and Consequences for Performance. Mis Quarterly 37(4), 1165-1188 (2013)
7. Dumont, L., Chamard, É., Léger, P.-M., Ortiz de Guinea, A. Sénécal, S.: Cognitive Analysis Grid for IS Research. In: Gmunded Retreat on NeuroIS (2014)
8. Baddeley, A.D., Hitch, G.: Working Memory. Psychology of Learning and Motivation 8, 47-89 (1974)
9. Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D.: The Unity and Diversity of Executive Functions and their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. Cognitive Psychology 41(1), 49-100 (2000)
10. Kramer, J.H., Mungas, D., Possin, K.L., Rankin, K.P., Boxer, A.L., Rosen, H.J., Bostrom, A., Sinha, L., Berhel, A., Widmeyer, M.: NIH EXAMINER: Conceptualization and Development of an Executive Function Battery. Journal of the International Neuropsychological Society 20(01), 11-19 (2014)
11. Zekveld, A.A., Heslenfeld, D.J., Johnsrude, I.S., Versfeld, N.J., Kramer, S.E.: The Eye as a Window to the Listening Brain: Neural Correlates of Pupil Size as a Measure of Cognitive Listening Load. NeuroImage 101, 76-86 (2014)
12. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.: Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist 38(1), 63-71 (2003)
13. Dumont, L., Gagnon, R., ElMouderrib, S., Gagnon, B., Théoret, H.: Validation of the French translation of the NIH-EXAMINER. In Progress.
14. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviation, Space, and Environmental Medicine 78(Supplement 1), B231-B244 (2007)
15. Sammer, G., Blecker, C., Gebhardt, H., Bischoff, M., Stark, R., Morgen, K., Vaitl, D.: Relationship between Regional Hemodynamic Activity and Simultaneously Recorded EEG-theta Associated with Mental Arithmetic-induced Workload. Human Brain Mapping 28(8), 793-803 (2007)

Research directions for methodological improvement of the statistical analysis of electroencephalography data collected in NeuroIS

Marc Fredette, Élise Labonté-LeMoyne, Pierre-Majorique Léger, François Courtemanche, and Sylvain Sénécal

Tech[3]Lab, HEC Montréal, Montréal, Canada
{marc.fredette, elise.labonte-lemoyne, pierre-majorique.leger, francois.courtemanche,ss}@hec.ca

Abstract. This proposed research will study and improve the statistical methodology used with neurophysiological data collected from subjects using information systems (IS). This research thus aims to provide guidelines and propose new statistical models constructed explicitly for the analysis of electroencephalography (EEG) data in IS research, where the number of EEG trials is often limited to preserve the ecological validity of the experiment. Two new modeling strategies are proposed: first, we will model explicitly the correlation between repeated trials by finding appropriate correlation structures. Secondly, we will reduce the measurement's error by using explicitly the cyclic behavior of an electrical brain signal. These new models will then be taken into account to derive new formulas for sample size determination.

Keywords. Toeplitz Structure · Exponential Structure · Periodic Functions · Hierarchical Likelihood · Statistical analysis,

1. Introduction

Reflecting on the emergence of NeuroIS and the challenges this nascent field of research is facing, [1] postulated that the establishment of a rigorous and comprehensive research methodology is capital. To achieve this goal, they argued that six factors will be critical: reliability, validity, sensitivity, diagnosticity, objectivity, and intrusiveness of a measurement instrument. It is important to note that a thorough and precise statistical analysis plays an essential role in the first five of these factors.

This proposed research will study and improve the statistical methodology used with neurophysiological data collected from subjects using Information Systems (IS). This type of data is usually complex and therefore subject to various statistical problems like high

measurement error and complex dependency patterns amongst the observations. In neuroscience, these statistical problems can often be solved by using an experimental design with a large number of repetitions to reduce measurement errors. The data are then aggregated to forego the modelling of complex dependency patterns. Statistical models that can be used with aggregated data are relatively straightforward and these simpler models are thus frequently used in this field. For example, a research made by [2] estimated that a simple method like the analysis of variance/covariance was used in 76% of the articles in psychophysiological research. They also noted that other simple methods like T-tests and multiple regression/correlation were each used in more than 20% of the papers. However, as we will discuss later, designs based on the aggregation of a large number of trials are often inappropriate for NeuroIS research because this can compromise the ecological validity of the experiment. This ecological validity is of the upmost importance to preserve the generalizability of the results obtained [3]. Therefore, we state that methodological research projects like this one have to be initiated within the NeurosIS community to increase the likelihood of finding significant results given these design's constraints.

Therefore, we will propose and study new models for electroencephalography (EEG) data that are more appropriate when the number of trials/events observed is small or moderate. This research thus aims to provide guidelines and propose new statistical models constructed explicitly for the analysis of neurophysiological data in IS research. These new models will thus increase the chances of finding new results and gain a better understanding in this field.

This research stems from the various research projects of the Tech^{3}Lab, a lab that specializes in the use of multi-method approaches in NeuroIS for measuring user's experience. At the Tech^{3}Lab, we jointly collect and analyze electroencephalography, along with eye-tracking and devices collecting many physiological measures (respiration, heart rate, galvanic skin response and facial emotions) while one or many subjects are interacting with a computer to perform a specific task (visiting a website, doing work-related tasks on a software, playing video games, etc.). A large amount of neurophysiological data has already been collected. For example, data collected in this lab were used in [4,5,6,7,8,9]. We will use these various datasets, and more datasets recently collected or to be collected soon, to implement and compare the models already used and the new ones proposed in this research. This will allow us to find which models are the most adequate (i.e., the ones providing the best goodness-of-fit) for typical IS research studies.

2. Efficient modelling of EEG data with moderate number of trials

In many research projects, we analyze a precise brain reaction in the seconds or milliseconds following a trigger of interest. However, in those event-related potential (ERP) or eye fixation-related potential (EFRP) experiments, the noisy signals coming from EEG data are difficult to analyze. In traditional neuroscience research, this problem is usually solved by collecting data on a very large number of ERP/EFRP trials in order to average out the noise. However, this is often not possible for research problems where we want to preserve the realism of an experiment. For example, if we want to study, throughout an EFRP experiment, the distraction caused by email pop-ups and the time spent to re-engage into the original task [7], the realism would be severely affected if hundreds of trials have to be made to reduce the noise. The need to develop efficient statistical methods to explicitly test EEG data with a limited number of ERP/EFRP trials in IS research is thus clear. We intend to investigate two different approaches to achieve this: First, we will take into account the correlation between repeated trials and find the best correlation structure to estimate it. Secondly, we will try to reduce the measurement error by using explicitly the cyclic behavior of an electrical brain signal.

2.1 Finding appropriate correlation structure for repeated EEG trials

Let us consider a typical scenario where a measure of interest is taken repeatedly after each trial for a subject. For example, after each auditory trigger, a Fast Fourier Transform was used to calculate the total power in the lower-frequency power in the alpha band in [6]. Very often, these measures are then averaged to reduce the measurement error. [10] confirms this view by saying, “if we average more and more observations, each with its own random error source but measuring the same true score, then the odds of the error canceling out keep improving … provid[ing] a more reliable measure” (p. 792). Nevertheless, aggregating a measure is less efficient than preserving all the individual measures and jointly analyzing them through a longitudinal regression approach [11, Section 3.6], as long as we take into account the potential correlation between all the repeated measures. Also as stated by [12]: "From a mathematical point of view, the basic problem is that complex functional relationships between two high-dimensional and highly variable signals (EEG and behavior) cannot be well characterized by first reducing each signal to a few average measures and then comparing them. Rather, what is needed is a new and quite different approach incorporating better recording and modeling of relationships between high-density EEG and more natural and higher-fidelity behavioral recordings."

Using various neurophysiological data collected in IS research studies previously mentioned, as well as many simulation studies using parameters similar to those observed in these studies, we want to find which covariance structure between repeated trials

provides the best fit to these types of data. A priori, the covariance structures that we intend to investigate are the compound symmetry structure (assuming a constant correlation among all trials), the Toeplitz structure (assuming that any pairs of responses that are equally separated in time have the same correlation), or a more general Exponential structure when the trials are not equally spaced over time [11, p. 173]. Once a suitable covariance structure is found, we will perform simulation studies to determine the threshold for the number of trials above which the loss of efficiency of using an aggregated approach is negligible.

2.2 Using cyclic behavior to reduce the measurement error of EEG data

Let us consider a frequent scenario where an ERP or EFRP experiment is conducted in order to study evoked potential after each trial. For example, this evoked potential is studied every time an email pop-up appears while someone is doing another task in [7]. It is well-known that the electrical signal in each part of the brain is made from a mixture of rhythmic and non-rhythmic activity. Once again, standard statistical analysis for these types of experiments are performed by calculating the mean amplitude of the average of all brain signals. When the number of trials is limited, we intend to investigate the added efficiency of explicitly modelling the rhythmic activity over a large number of plausible frequencies in order to reduce the measurement error.

Let t_{ij} be the time the jth trigger is activated for the ith subject where i=1,..,N and j=1,…,n_i and let $v_i(t)$ be the observed voltage for the ith subject at time t at a given electrode. Let us assume, without any loss of generality, that each subject is assigned by design to one and only one group. Usually, the mean amplitude over a given time interval [a,b] (corresponding to the latency) for a given subject is estimated by approximating the integrals:

$$V_{i,(a,b)} = \frac{1}{n_i}\sum_{j=1}^{n_i}\int_{t_{ij}+a}^{t_{ij}+b} v_i(t)dt.$$

These estimated values for each subject are then usually compared using ANOVAs or T-tests according to the number of groups tested. With such statistical models, the (normally distributed) error term can be decomposed into three parts: a subject-specific effect, the rhythmic activity observed during the interval [a,b], and the true measurement's error. With a limited number of trials for each subject, it is plausible that the error due to the rhythmic activity of the brain will account for a large part of the error term. Therefore, we propose to explicitly model the rhythmic activity of the brain by using the following model:

$$v_i(t) = \mu_i(t) + \sum_{p=1}^{P} a_p f_p(t) + \varepsilon_i(t),$$

where $\mu_i(t)$ is the non-rhythmic activity that we want to study, $\varepsilon_i(t)$ is the error term, the a_p's are P unknown parameters to be estimated, $f_p(t)$ is a periodic function having a period of w_p, and $(w_1, \dots, w_P)$ is the set all considered periods to model the brain rhythmic activity for a given electrode. If the periods and the periodic functions used are close enough to the true rhythmic activity of the brain, this model will greatly reduce the number of trials needed to obtain an ideal "signal to noise ratio", a measure reflecting the ability to distinguish signals from noise [13]. An important aspect of this research will thus be to use datasets observed from typical NeuroIS studies to determine the right number of periods and the most appropriate periodic function (which is not necessarily a sinusoidal function).

3. Concluding remarks and future work

We feel that the research directions stated in this paper should initiate more statistical research projects in NeuroIS. Our community would greatly beneficiates of using more efficient statistical models when the number of EEG trials has to be limited to preserve the ecological validity of an experiment.

A subsequent long term research objective is to develop formulas, based on the models found above, to determine the appropriate sample sizes for research studies in NeuroIS. Sample size determination is an important concept in this field because data collections are often expensive. It is thus valuable to find the number of subjects needed to achieve a desired power before one commits the resources. Developing an algorithm to find the right number of subjects and the right number of trials is more difficult than usual methods because of the presence of repeated measures amongst each subject. We intend to study two promising avenues: First, we can extend the methodology developed by [14] for fMRI (functional magnetic resonance imaging) studies. Another possibility is to extend the general approach of [15] who proposed to make calculations based on a hierarchical likelihood, called the H-likelihood.

References

1. Riedl, R., Davis, F.D., Hevner, A.R.: Towards a NeuroIS Research Methodology: Intensifying the Discussion on Methods, Tools, and Measurement. Journal of the Association for Information Systems 15(3), Article 4 (2014)
2. Jennings, J.R., Gianaros, P.J.: Methodology. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (eds.) Handbook of Psychophysiology (3rd ed.), pp. 812-833. Cambridge University Press, Cambrige (2007)
3. Lee, A.S., Baskerville, R.L.: Generalizing Generalizability in Information Systems Research. Information Systems Research 14(3), 221-243 (2003)

4. Mirhoseini, S., Léger, P.M., Sénécal, S., Fredette, M., Cameron, A.F., Riedl, R.: Investigating the Effect of Input Device on Memory Performance: Evidence from Theta and Alpha Band Oscillations. In: Proceedings of the SIG-HCI Conference, Paper 15. AIS (2013)
5. Sénécal, S., Léger, P.M., Fredette, M., Courtemanche, F., Cameron, A-F., Mirhoseini, S., Paquet, A., Riedl, R.: Mouse vs. Touch Screen as Input Device: Does it Influence Memory Retrieval? In: Proceedings of the ICIS 2013 Conference. AIS (2013)
6. Courtemanche, F., Léger, P.M., Cameron, A.F., Faubert, J., Labonté-LeMoyne, É., Sénécal, S., Fredette, M., Bellavance, F.: Texting While Walking: Measuring the Impact on Pedestrian Visual Attention. In: Proceedings of the Gmunden Retreat on NeuroIS 2014 (2014)
7. Léger, P.M., Sénécal, S., Courtemanche, F., Ortiz de Guinea, A., Titah, R., Fredette, M., Lemoyne, É.L.: Applying Eye Fixation-Related Potentials to Information Systems Research: Demonstration of the Method during Natural IS Use and Guidelines for Research. Journal of the Association for Information Systems 15, 651-678 (2014)
8. Labonté-LeMoyne, É., Santhanam, R., Léger, P.M., Courtemanche, F., Fredette, M., Sénécal, S.: The Delayed Effect of Treadmill Desk Usage on Recall and Attention. Computers in Human Behavior 46(1), 1-5 (2015)
9. Sénécal, S., Fredette, M., Léger, P.M., Courtemanche, F., Riedl, R.: Consumers' Cognitive Lock-in on Websites: Evidence from a Neurophysiological Study. To appear in Journal of Internet Commerce (2015)
10. Strube, M. J., Newman L.C.: Psychometrics. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (eds.) Handbook of Psychophysiology (3rd ed.), pp. 789-811. Cambridge University Press, Cambridge (2007)
11. Fitzmaurice, G.M., Laird, N.M., Ware, J.H.: Applied Longitudinal Analysis (2nd Edition). Wiley, Hoboken, New Jersey (2011)
12. Makeig, S., Gramann, K., Jung, T.P., Sejnowski, T.J., Poizner, H.: Linking Brain, Mind and Behavior. International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology 73(2), 95-100 (2009)
13. Gratton, G.: Biosignal Processing. In: Cacioppo, J.T., Tassinary, L.G., Berntson, G.G. (eds.) Handbook of Psychophysiology (3rd Edition), pp. 834-858. Cambridge University Press, Cambridge (2007)
14. Mumford, J.A., Nichols, T.E.: Power Calculation for Group fMRI Studies Accounting for Arbitrary Design and Temporal Autocorrelation. Neuroimage, 39, 261-268 (2008)
15. Lee, Y., Nelder, J.A.: Hierarchical Generalized Linear Models. Journal of the Royal Statistical Society - Series B (Methodological) 58(4), 619-678 (1996)

Measuring visual complexity using neurophysiological data

Vanessa Georges[1], François Courtemanche[1], Sylvain Sénécal[1], Thierry Baccino[2], Pierre-Majorique Léger[1] and Marc Frédette[1]

[1] Tech3Lab, HEC Montréal, Montréal, Canada
{vanessa.georges, francois.courtemanche, sylvain.senecal, pml, marc.fredette}@hec.ca
[2] LUTIN User Lab, Université de Paris 8, Paris, France
baccino@lutin-userlab.fr

Corresponding author: vanessa.georges@hec.ca

Abstract. The effects of design and aesthetics on interface usability has become an important research topic in recent years. In this paper, we propose a new method of interface visual complexity evaluation based on the users' neurophysiological signals. In order to be truly insightful, a visual representation of such signals will be mapped onto the interface using physiological heatmaps. The method's intended purpose is to inform practitioners and researchers in information system on how different interface designs affect perceived visual complexity.

Keywords: User Experience · visual complexity · eyetracking · heatmaps · interface design · neurophysiological signals

1 Introduction

User experience (UX) has recently become of strategic importance in the information technology industry. UX is defined as a person's perceptions and responses that result from the use or anticipated use of an IT product or service [1]. Large software companies such as SAP are now primarily positioning their products on the user experience, and specifically in the simplicity of their user interface (e.g., http://discover.sap.com/runsimple).

Researchers suggest that UX is composed of three main dimensions: pragmatic (focuses on the IS usability), emotional (focuses on emotional responses triggered by the IS interaction), and hedonic (focuses on the visual, symbolic, and motivational features of the IS) [2-4]. The latter dimension remains the lesser child of the overall UX. This paper proposes a neurophysiological method of UX evaluation measuring a key construct of the hedonic dimension: visual complexity [5].

2 Visual complexity

A review of the literature identified visual complexity as a key concept in predicting users' aesthetic appeal [6,7]. According to Oliva *et al.* 2004 [8], visual complexity is defined by "*the degree of difficulty in providing a verbal description of an image*". The relationship between visual complexity and affective valence follows an inverted U-shaped curve [9]. Interfaces at both extremes of the curves, either deemed too simple or too complex, will result in lower affective valence.

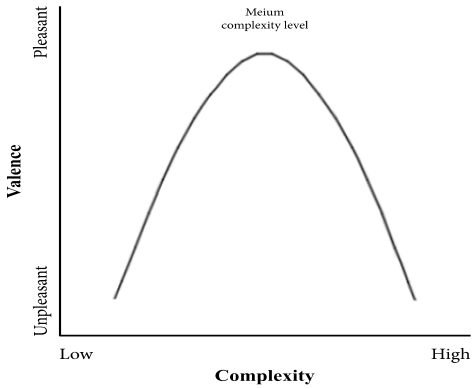

Fig. 1. The relationship of visual complexity and affective valence.

In recent years, evaluation methods of visual complexity have been based mainly on image characteristics (e.g. number of pixels, frequency), user performance (response time), design heuristics, and HTML code. However, no evaluation method has focused on the actual IS user. Therefore, we propose a method of evaluating visual complexity directly relying on user experience, using neurophysiological signals.

3 Proposed Method

The proposed method builds upon our previous work proposing a visual representation technique based on the triangulation of eyetracking and physiological data called Physiological Heatmaps [10]. These heatmaps are a novel visualization method which represents the relative intensity of a physiologically inferred affective or cognitive state on an interface using heatmaps' color gradients (Figure 2). Physiological heatmaps can be adapted to different psychological constructs (e.g., discrete emotion, cognitive load) by training the inference engine (machine learning model) on a related data set.

The objective of this research is to develop a method to assess the visual complexity of the different elements of an interface. Physiological heatmaps will be used to map visual complexity onto the evaluated interface. In this work an experiment will

be conducted to produce a training data set of physiological signals related to different level of perceived visual complexity (see section 4).

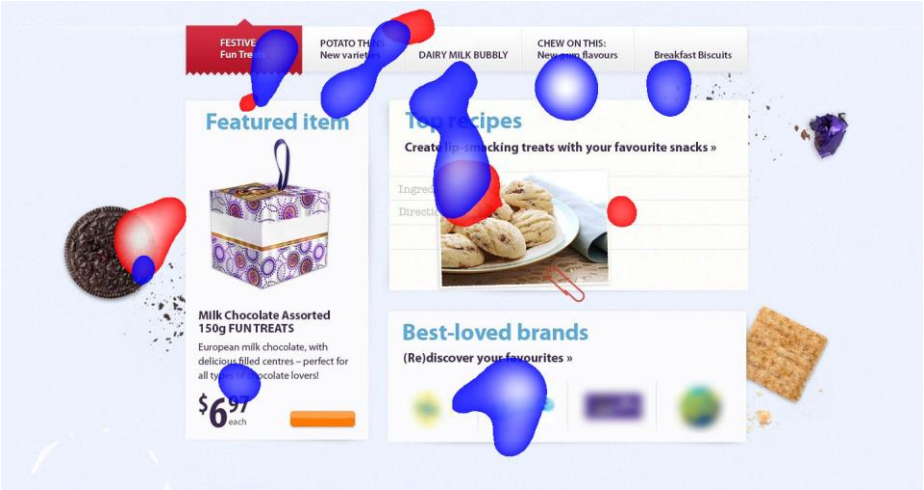

Fig. 2. The blue gradient represents a standard gaze heatmap and the red gradient represents an arousal physiological heatmap based on pupil size.

Once validated, the proposed method will provide UX and IS researchers and practitioners with a new tool to better inform decisions during the design development of a computer interface, at various stages of prototyping. In other terms, as the development of the interface progresses, professionals will be able to keep track of how design changes affect users' visual complexity perceptions, and therefore better guide their decisions. For example, it will allow comparing the experienced visual complexity of two versions of an interface (A/B testing).

4 Experiment

The experiment carried on in this work will provide a physiological data set allowing the training of the visual complexity inference engine and the evaluation of the proposed method. Images from the homepage of 24 websites will be used as stimuli. These interfaces will include three visual complexity conditions: low, medium and high (evaluated by experts). As illustrated in Figure 3, the experiment will consist of six presentation blocks (two per condition), each containing four stimuli. Each block will be separated by a vanilla baseline period [11], and intra and inter block orders will be randomized.

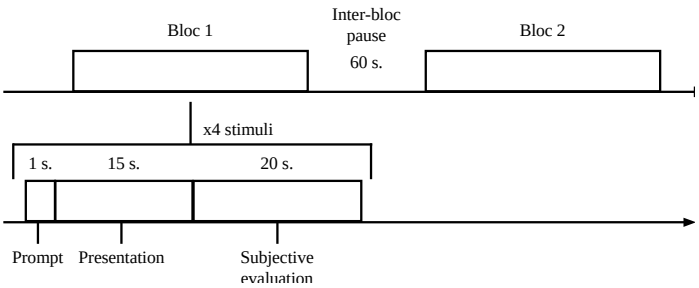

Fig. 3. Four images will be presented within each condition block. Blocks will be separated by a 60 seconds rest period to ensure the return of physiological signals to baseline level.

The experiment will take place during the month of May 2015. A total of 30 volunteer students between the ages of 18 and 35 will be recruited through HEC Montréal's student panel. After each stimulus presentation, the subjective evaluation will consist of the following steps:

- Participants will be asked to indicate by a mouse click the areas of the interface which are visually more complex to them.
- Participants will then be asked to rate these areas on a scale of one to ten.
- Participants will be asked to complete a post-experimental questionnaire in order to assess their overall appreciation of the interface.

Research has established a strong relation between experienced visual complexity and cognitive load [12, 13]. As stated by Harper *et al.* 2009 [13] (p. 14), "*Visual complexity seems to be an implicit key into the perceived cognitive load of the page and the interaction that the users think will be required to use the resource. As such, we can use an analysis of the visual complexity to give us an approximation of the cognitive interaction load required by the page.*" Therefore, the inference engine underlying visual complexity heatmaps will use peripheral physiological signals related to cognitive load, such as pupil size [14], heart rate [15], and electrodermal activity [16]. Emotional valence will also be measured using the FaceReader 6 (Noldus, The Netherlands) facial expressions analysis software. Validation will consist of correlation analyses between interfaces' visual complexity measured by physiological heatmaps and by users' subjective evaluations.

5 Conclusion

In this paper, we proposed a new method for interface visual complexity evaluation based on physiological heatmaps. The method will guide UX and IS practitioners and researchers in the various stages of interface development. This novel way of exploring visual complexity will lead to a better understanding and definition of the hedonic dimension of UX while interacting with computer interfaces.

References

1. ISO, W., 9241-11. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs). The International Organization for Standardization (1998)
2. Février, F., Gauducheau, N., Jamet, E., Rouxel, G., Salembier, P.: La prise en compte des affects dans le domaine des interactions humain-machine. Le Travail Humain 74, 183-201 (2011)
3. Mahlke, S. Minge, M.: Consideration of Multiple Components of Emotions in Human-Technology Interaction. In: Peter, C., Beale, C (eds.) Affect and Emotion in Human-Computer Interaction Editors, pp. 51-62. Springer, Berlin, Heidelberg (2008)
4. Hassenzahl, M.: The Interplay of Beauty, Goodness, and Usability in Interactive Products. Human–Computer Interaction 19(4), 319-349 (2004)
5. Tuch, A.N., Bargas-Avila, J.A., Opwis, K., Wilhelm, F.H.: Visual Complexity of Websites: Effects on Users' Experience, Physiology, Performance, and Memory. International Journal of Human-Computer Studies 67(9), 703-715 (2009)
6. Michailidou, E., Harper, S., Bechhofer, S.: Visual Complexity and Aesthetic Perception of Web Pages. In: Proceedings of the 26th Annual ACM International Conference on Design of Communication. ACM (2008)
7. Ngo, D.C.L., Teo, L.S., Byrne, J.G.: Modelling Interface Aesthetics. Information Sciences 152, 25-46 (2003)
8. Oliva, A., Mack, M.L., Shrestha, M., Peeper, A.: Identifying the Perceptual Dimensions of Visual Complexity of Scenes. In: 26th Annual Cognitive Science Society. (2004)
9. Berlyne, D.: Studies in the New Experimental Aesthetics. Hemisphere Publishing, Washington, DC (1974)
10. Courtemanche, F., Léger, P.-M., Frédette, M., Sénécal, S., Dufresne, A.: Method and Product for Visualizing the Emotions of a User. In: Provisional Patent Application, p. 14. (2015)
11. Jennings, J.R., Kamarck, T., Stewart, C., Eddy, M., Johnson, P.: Alternate Cardiovascular Baseline Assessment Techniques: Vanilla or Resting Baseline. Psychophysiology 29(6), 742-750 (1992)
12. Wang, Q., Yang, S., Liu, M., Cao, Z., Ma, Q.: An Eye-tracking Study of Website Complexity from Cognitive Load Perspective. Decision Support Systems 62, 1-10 (2014)
13. Harper, S., Michailidou, E., Stevens, R.: Toward a Definition of Visual Complexity as an Implicit Measure of Cognitive Load. ACM Transactions on Applied Perception (TAP) 6(2), 10 (2009)
14. Laeng, B., Sirois, S., Gredebäck, G.: Pupillometry. Perspectives on Psychological Science 7(1), 18-27 (2012)
15. Task Force of the European Society of Cardiology the North American Society of Pacing: Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. Circulation 93(5), 1043-1065 (1996)
16. Parasuraman, R.: Neuroergonomics Brain, Cognition, and Performance at Work. Current Directions in Psychological Science 20(3), 181-186 (2011)

Using NeuroIS to Better Understand Activities Performed on Mobile Devices

Carole L. Hollingsworth[1] and Adriane B. Randolph[1, (✉)]

[1] Kennesaw State University, Department of Information Systems, Atlanta, Georgia, USA
{chollin2,arandolph}@kennesaw.edu

Abstract. With the proliferation of mobile device types and variety of tasks being performed on those devices, it is necessary to examine how this pairing changes with individuals. NeuroIS offers complementary tools to traditional survey tools helping researchers delve into users' perceptions while they are engaged in different tasks. Through analysis of neurophysiological data we may better understand activities performed on mobile devices and help provide more customized user experiences. A two-part preliminary study is described as a pre-cursor to a larger, focused experiment utilizing EEG and Eye Tracking on mobile device usage.

Keywords: Mobile device usage · hedonic task · utilitarian task · eye-tracking · EEG · individual differences · NeuroIS

1 Introduction

Presently, there are nearly 7 billion mobile phone subscriptions in use throughout the world [1]. These devices range from the highly personal smartphone to the tablet-based computer which may be shared among users to the intimate wearable devices which are now frequently entering the category. Users interact with their mobile devices in different and engaging ways including customization far beyond the software itself or the cases that are used to protect them. This customization extends to how, when and where users choose to employ them. Smartphones and tablets are often on, accessible to the Internet, are typically with a user wherever they go and have actively become part of an individual's daily routine. Thus, it is helpful to examine the activities that individuals are willing to complete on these tools.

Learning how users interact with mobile devices will allow for better development of technologies and also development of applications which are best tailored to user preferences. Leveraging the tools available to NeuroIS researchers, it is possible to gain a deeper understanding of users' perceptions while performing these tasks. Although the devices themselves seem to be created with a 'one device fits all users' approach, there may be significant differences in what different groups of users are willing to do on the specific devices. Here, we present a multi-stage study seeking to

uncover these differences using neurophysiological recordings of user interactions with a smartphone and tablet across varying task types.

2 User Tasks and Mobile Devices

2.1 Hedonic, Utilitarian and Mixed Activities

Activities performed on mobile devices typically can be classified in one of three ways: hedonic, utilitarian or mixed, where mixed is a combination of a hedonic and utilitarian activity. Van der Heijden [2] examined the contrasts between user acceptance of hedonic information systems (pleasure-oriented) and utilitarian information systems (productivity-oriented) yet mobile devices blur the lines between them. Likewise, hedonic motivation to use an information system has recently been explored from the perspective of the consumer [3] and seen to be likely that consumer preferences and use will differ based on devices, technological capabilities, other tastes and even age. One can argue that the simple act of using a mobile device is somewhat enjoyable as the experience uses non-traditional input means; users touch the device with their fingers or a stylus or they might speak to the device. There are hand gestures, which although simple, make the user more a part of the human-computer interaction. Likewise, the device is often held in one or both hands, may or may not be placed on a table or it might even be adorning the user as is the case with wearables. Thus, the connection becomes more intimate and personal between the user and the mobile device at the time of interaction.

A key component of a hedonic information system would be that of enjoyment. Perceived enjoyment has been shown to have an impact on the use of a system [2,4,5]. This would be important if an ordinarily less enjoyable task is perceived to be enjoyable while using a mobile device. Additionally, a hedonic system, or one that is fun to use offers value in the interaction between system and user, and a utilitarian system offers value outside the interaction of user and system, such as increased productivity [6].

Answering the call to extend NeuroIS research [7] into a mobile technology focused study, this work seeks to develop a protocol which can examine users while performing hedonic, utilitarian and/or mixed activities. A typical hedonic task can include posting to social networking sites, engaging in online shopping or playing a computer game. Utilitarian tasks could include using an ERP system, using a university's learning management system or similar application. Previously, mixed activities have had elements of both hedonic and utilitarian tasks. A good example of a mixed task is using e-mail. This has shifted over the years as personal e-mail might be enjoyable while work e-mail might be more functional or utilitarian.

2.2 Focused Activities

For the purposes of this study, only a hedonic activity and a utilitarian activity will be examined as a mixed activity may offer some overlapping and the intent is to have

differentiation between types of activities. Instead, these two activities will be examined on several mobile device types. To simplify and control against any user confusion, one operating system will be selected. For example, instead of performing a task on an iPhone and then on a Samsung Galaxy tablet, Apple's iOS or an Android OS will be selected for testing purposes. The preliminary study uses Apple devices and iOS version 8.1.3.

Participants will complete two tasks, each one on an Apple iPhone 5s and an iPad Air. For the utilitarian task, the participant will be enrolled in the university's learning management system. The user will take a one-question, fill-in-the-blank quiz on each device. Students are the targeted population for participants in this activity since they are knowledgeable about the nuances of the learning management system and are familiar with the devices being tested. For the hedonic task, the devices are preset and logged into an active social media account which belongs to the sponsoring research center. The user will then type a sentence into the account preparing to make a social media posting. Since the utilitarian activity is simulating a test experience, the social media posting is also simulated and is typed in but not posted.

3 Proposed Study & Protocol

A recent survey-based study at a large comprehensive university in the Southeastern United States examining user preferences for performing specific tasks on different mobile devices yielded a desire to better understand the habits and preferences of this population as users of the technologies. To build a protocol for a larger focused experimental study, there is a two-part preliminary study being completed in advance. Once the preliminary study is complete, the final protocol will be further refined for the larger focused experiment.

3.1 Preliminary Study - Part 1 – Eye Tracking

Utilizing a Tobii Glasses 1 Eye Tracker, the first part of the preliminary study concentrated on gaining an understanding of what the user is examining while performing each of the activities. To use the system, each individual user has to be calibrated to the system to ensure it can view and record his/her eye movement. The device uses two cameras with one trained on the user's gaze and a second on the user's eye. Once appropriately calibrated, with the system confirming adequate accuracy and tracking ability, recordings can be made and saved. Accuracy and tracking ability is measured in levels of up to five stars, with accuracy being the more important measure.

For this preliminary study, accuracy reached five stars in three activities and four stars in the other. Four separate recordings from the glasses were examined and analyzed through the Tobii Studio Eye Tracking software. Within the software, a video showed what areas a user was viewing and then overlayed a dot and vector mapping. The large dots are areas where the focus has been for longer than one second. Lines then demonstrated the eye movement and pathway followed. Following in Figures 1 and 2 are screenshots from a participant executing a hedonic task on a smartphone and tablet.

Examining each of the activities, the user, who actively uses these technologies had no complications performing the tasks. Since the participant already uses the technology, it was fully anticipated that the results would demonstrate comfort and proficiency

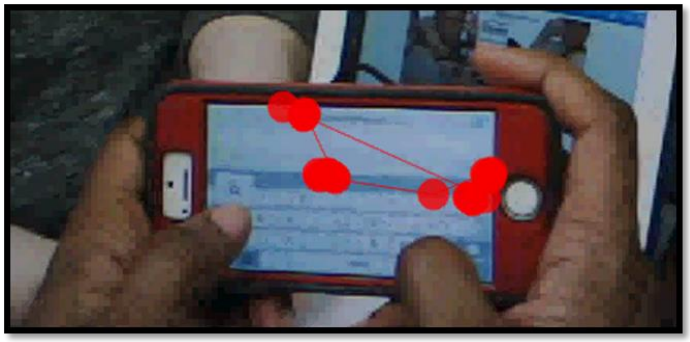

Fig. 1. Tobii Studio Software Showing Eye-Gaze for Hedonic Smartphone Task

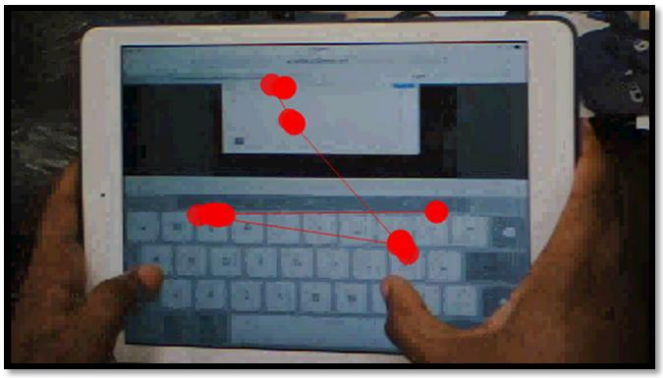

Fig. 2. Tobii Studio Software Showing Eye-Gaze for Hedonic Tablet Task

There were some interesting findings when examining the videos. The participant is female, age 37 and uses corrective contact lenses but is able to see the devices without issue. In both devices, the predictive text systems which are part of Apple's iOS8 were active as is common when using either device while typing. This was left on in all activities as many users do take advantage of the capabilities. First, the two hedonic activities were examined against each other and then the two utilitarian activities were examined against each over. Both yielded a common thread which was that the participant tended to use the suggested words whenever there was an option when using the smartphone but did not do so when using the tablet. When asked after the activity, the participant indicated, it was easier viewing the intended text to type on a tablet than a smartphone and then she did not rely on the predictive text. Also, due to the smaller keyboard of the smartphone, the predictive text system was helpful. Differences also occurred in that it

took longer to complete the hedonic activity on the smartphone than it did on the tablet. It took approximately a third less time on the tablet. Likewise, completing the utilitarian activity also took approximately one third less time on the tablet versus the smartphone. Perhaps the participant's comfort with a larger device might suggest that she consider moving to a larger smartphone to gain more efficiency. It will be interesting to see if this holds in the larger focused experiment. Future opportunities can examine different age groups and populations based on their device use.

3.2 Preliminary Study – Part 2 - EEG

The next step, is to record EEG while performing the same tasks. Sixteen channels of EEG will be recorded using the BioSemi Active Two bioamplifier system connected to a Windows-based computer [8]. An electrode cap will be fitted according to the frequently used best practice of the 10-20 system of electrode placements [9]. The electrodes will be placed on the cap to permit recording of brain activations over the frontal lobe and scalp and will be sampled at 16384 Hz using a Common Average Reference (CAR). The sixteen channels recorded will be: Fp2, Fp1, F4, Fz, F3, T7, C3, Cz, C4, T8, P4, Pz, P3, O1, Oz, O2 – where electrodes starting with the letter F cover the frontal and pre-frontal (Fp) lobe. Following the activities, each of the four separate recordings from the sixteen channels of scalp-based electrodes will be analyzed offline using a previously validated technique for brain localization and associated software, standardized low resolution brain electromagnetic tomography (sLORETA) [10].

3.3 Focused Experiment Protocol – EEG & Eye Tracking

For the focused experiment, both EEG and Eye Tracking will be employed using the previously discussed methods. For the Eye Tracking portion, participants who wear contact lenses regularly should be tested in advance to ensure calibration levels can be achieved. Following each participant, the EEG and Eye Tracking recordings will be evaluated as previously discussed. If a second operating system is included, then there will be further comparison between operating systems to see if there are any differences at the user level.

4 Conclusion & Contribution

This research is still in progress yet there are hopes to better inform the field about the types of activities which are best suited to specific mobile devices by complementing traditional survey methods with neurophysiological tools. Using the combination of EEG and Eye Tracking will offer a different perspective both of the user and by the user when completing different activities. With the proliferation of mobile devices and continued use, a better understanding will help make the device use be productive and effective and not a nuisance. Future research will be focused on examining different age groups, users with differing capabilities and different mobile devices.

References

1. International Telecommunications Union. (2014). ICT Facts and Figures 2014. Retrieved from: http://www.itu.int/ITU-D/ict/facts/material/ICTFactsFigures2014.pdf
2. Van der Heijden, H.: User Acceptance of Hedonic Information Systems. MIS Quarterly 28(4), 695-704 (2004)
3. Venkatesh, V., Thong, J., Xu, X.: Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 36(1), 157-178 (2012)
4. Venkatesh, V.: Creation of Favorable User Perceptions: Exploring the Role of Intrinsic Motivation. MIS Quarterly 23(2), 239-260 (1999)
5. Venkatesh, V.: Determinants of Perceived Ease of Use: Integrating Perceived Behavioral Control, Computer Anxiety and Enjoyment into the Technology Acceptance Model. Information Systems Research 11(4), 342-365 (2000)
6. Gerow, J.E., Ayyagari, R., Thatcher, J.B., Roth, P.L.: Can We Have Fun@ Work? The Role of Intrinsic Motivation for Utilitarian Systems. European Journal of Information Systems 22(3), 360-380 (2013)
7. Dimoka, A., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P.H., Pavlou, P.A., Müller-Putz, G., Riedl, R., Vom Brocke, J., Weber, B.: On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. MIS Quarterly 36, 679–702 (2012)
8. Active Two. Retrieved from http://www.cortechsolutions.com/Applications/Brain-Computer-Interface/ActiveTwo
9. Homan, R.W., Herman, J., Purdy, P.: Cerebral Location of International 10–20 System Electrode Placement. Electroencephalography and Clinical Neurophysiology 66(4), 376-382 (1987)
10. Pascual-Marqui, R.D.: Standardized Low-resolution Brain Electromagnetic Tomography (sLORETA): Technical Details. Methods and Findings in Experimental and Clinical Pharmacology 24 (Suppl D), 5-12 (2002)