Using fNIRS to Detect Mental Workload and Emotional Valence in web form filling task

Kristiyan Lukanov

2015-09-18

Acknowledgements

blagodarq na toq i onq iuawhdauiwdaw joaiw joiakj aoif oiak opaekgpo kaeoikae

Abstract

taka i taka oijga geriugj eurgjergijeigire eg jg eogpergkeprkgoe kpokepokgekgkeg-pokegpokepgkegk tgea tgwe gwe ghwgwhgrjtujdty

Contents

Chapter 1

Introduction

- Start with the problem

Users often has to fill web pages containing more than 10 forms, like registering for a web site, posting classified ad, or sending online insurance claim. Sometimes this is really important, like filling insurance claim forms. People often dislike the fact that they should fill a long form and that makes the claim process difficult. That is why we are interested in measuring the mental workload in this type of tasks

we want to if new web form designs will make the task easier

Importance of those forms(how they should be accurate and unaiding) However this often encountered task in our daily lives is not well researched in the field of cognitive science.

-it is important measure, especially for critical tasks such as, aircraft controller, where operator with workload overload could cause accidents and cost human lifes.

With the advance of technology, brain imaging techniques have emerged, like MRI, EEG, and fNIRS which were initially, used for medical research and purposes. In recent years the popularity of these brain imaging techniques has risen in the HCI field because of their ability to image brain data.

videos shown are aversive stimuly(ili otbluskva6ti stimuli, ne e sigurno da se proveri) .

The studies using brain imaging were limited because they were presenting emotional cues, but in this study we check whether we can elucidate laterized emotions giving subjected emotional cue-less web interface. hich will be very useful in HCI area

1.1   Purpose of study

The purpose of the study is to find whether fNIRS can be useful for interface evaluations in HCI field. Also, we aim to find the most efficient layout for entering information on an web interface that has more than 10 forms , as this process is often encountered during daily web surfing, for example, when user registers to a new web site, or enter information for financial institutions, like insurance companies and banks.

1.2 Research questions

Can we measure mental workload with fnirs? - Could we detect emotional valence with fnirs? - Which of the three layouts is has the least mental workload and users preference? - Could we detect emotional valence, from web interface that has no emotional cues. This can very useful in HCI evaluations.

1.3 Industry partner

This work has been motivated by the need of entity partner funding my masters course. The industry partner has a insurance CRM software, and the aim of the study is to provide insights in the web form filling process by testing the layout of the web pages.

1.4 Structure of the thesis

In the next chapter we will first review the background literature behind usability and web forms, mental workload and working memory, Brain sensing techniques, Emotion processing. In chapter 3 we will describe the User study section experiment done in the . And finally we will discuss the results from the experiment and propose a conclusion.

Chapter 2

Literature review

In the HCI field evaluation approaches ca be divided in three general categories: analytical, field study and lab study[36]. First, analytical methods are designed to predict user behaviour such as, heuristic evaluation or expert reviews, so no experiment has to be conducted. Second, field studies are conducted in context in order to collect relevant and valid data, like observations. Lastly, lab studies use artificial settings but the experiment variables can be controlled easier, and also, comparative tests can be conducted. Our aim of the study is to evaluate a web form filling interface for the insurance domain, and more precisely, online auto insurance claim process. Therefore we are interested in conducting lab study because we cannot simulate road accident and we are able to compare variations of insurance claim web form. Furthermore, there are variety of evaluation methods, like interviews which will give us information about what users think about the interface, or observations which will let us recognize typical behaviour of users and obstacles they encounter while using the interface.

It has to be mentioned that lab studies give us the chance to prepare the environment and record more performance measures which provide valuable objective information. Typical performance measures in usability experiments are time to complete, errors encountered, and number of events. In general, mental workload and emotional valence are of high interest in HCI evaluations because measuring workload gives us important information about the task demands and also, knowing whether a user is feeling positive or negative towards an interface or task may give us valuable information about their general preferences. Furthermore, Nielsen[28] suggested that user preferences correlate with user performance, thus we can rely solely on user preferences, however this is not always the case. Hence, Nielsen and Levy [28] advise researchers to use combination of subjective and objective data in usability studies, in order to identify bias and provide richer information about the process. Accordingly, we have decided to employ user trials(subjective data) combined with psychophysiological measurements(objective data). In addition, functional near infrared spectroscopy (fNIRS), has been recently suggested as a promising method for HCI evaluations[23, 34] because it was suggested to measure mental workload[24]. Based on this, we decided to test the usefulness of fNIRS in HCI evaluation studies.

Based on the information above this master thesis considers measuring mental workload and emotional valence to inform the user interface design of the insurance claim web form. The literature review will proceed in the following way: first, we will review relevant literature on web form filling and usability. Second, we discuss working memory models and the mental workload concept. Third, emotional processing literature will be revised. Fourth, current brain sensing techniques used in cognitive experiments will be reviewed and, fifht, fNIRS studies examining mental workload and emotional valence will be examined. Finally, we will summarize the reviewed literature.

2.1 Usability and Web form filling

Web form filling is often encountered activity in daily surfing of web users, however, according to our knowledge, there is scarce of empirical research in the Human Computer Interaction(HCI) literature for this topic. First, a study by Wstlund[47] compared two web page layouts - one that all the text is in the same page, and one where the text is separated in four pages. Authors concluded that users experienced less workload with the divided web form(4 pages), compared to the single page web form. Second, two books specially written for web form filling design[18, 46] suggest splitting long web forms into several pages, in order to improve the process. Lastly, most of the research on web form filling and design is focused in optimizing the experience and accessibility for elderly population[39, 10, 20, 38].

In reviewing positive and negative affect and their implications for design Norman[30] proposes that "Positively valenced affect broadens the thought processes hence, enhanced creativity". Therefore, increase in task performance should be observed when a user has positively valenced affect towards certain interface. Suggesting that emotional valence plays pivotal role in task performance. Also, A couple of studies suggest that the longer it takes for a task(short or long term) to be completed the more the perceived frustration the users experience increases[25, 4]. Therefore, we should aim to minimize frustration by reducing the time to complete a task.

Because of insufficiency of relevant literature on the web form filling and design we are going to examine general usability guidelines and recommendations as they are widely accepted by the Human Computer Interaction researchers. The two most popular usability heuristics are those of Nielsen[29], and Shneidermann[40]. They express similar suggestions, like, maintain consistency, provide feedback, support expert users, prevent and optimize error messages, provide help documentation, permit easy reversal of information, and minimize working memory load. Consequently, the design of the tested variants in the usability study in chapter 3 is informed by them. Also, because both heuristics advocate minimizing the load on working memory we consider that reducing it will provide better user experience. In addition, because usability of certain interface depends on the context, user differences, and that there is not perfect solution to a interface problem, and designers often have to make tradeoffs[32], we will rely on cognitive science in order, to predict which layout is more appropriate.

2.2 Working Memory and Mental Workload

The concept of mental workload (MW) is intuitive in nature and it represents how busy an operator is when performing a certain task. The concept has been referred in the literature with many terms, like cognitive load, stress, strain, and arousal. The aim is to Many definitions has been proposed by many authors, however researchers are still unable to find a consensus on the term[Linton et al 1989]. Wickens[43] defines it as "The demand imposed by tasks on the human's limited resources, whether considered single or multiple". Depending on the studied task at hand, knowing workload experienced by different design variations will help choose the one that generates desired operator performance. Also, in terms of operator experience of MW, Rouse et al classifies different factors like, fatigue, mood, individual differences, as person-specific workload[37]. Similarly, Norman and Bobrow classified operator performance on data-limited and resource-limited[31]. They hypothesize that even if operator spends high amount of attentional resources, the task can have a bad representation that will degrade the performance. In contrast, resource-limited performance depends on how much attentional resources the task demands, and it can be considered that every real life task consists of combination of both.

2.2.1 Working memory models

Rather than searching for definition researchers in cognitive science use models of working memory in order to understand cognition, predict and explain workload and performance. Furthermore, theories of working memory try to define the processes going into human mind, and explain concepts such as, attention, perception, long term memory, decision making, action selection, and execution[44, 2, 26]. Most of those models are based on human as information processor approach [6, 7, 27, 44], which relates the processes of human mind with those of a computer processor. Also, the framework is based on the assumption that the human operator has a limited resource capacity[], and if the task demands more resources than the capacity of the operator, workload overload is observed. Moreover, the information from the environment or the task is processed by series of processing systems, like perception, attention, short-term memory, long-term memory. In attempt to describe the web form filling task we can use the working memory model from Baddeley and Hitch [2] which processes information in verbal and spatial form . It consists of a central executive, which is acts as an administration system which controls the information input and output of its slave systems. The visuo-spatial sketch pad is involved in holding visual information in spatial form like, objects and colours. The phonological loop stores verbal information, such as words and names. And the later proposed [1] episodic buffer is responsible for the storage and retrieval of memories or events. Because the task of web form filling involves multiple cognitive processes like, visual search, speech synthesis, planning, memory retrieval, decision making, thus utilizing all slave systems of the model, we can label the web form filling process as one that involves complex cognition.

We can also consider the multiple resources model by Wickens[43, 42] which is suited for predicting the workload of an operator performing multiple tasks at one time. The approach is based on four basic assumptions:

Figure 2.1: Working memory model by Baddeley and Hitch, displaying the 'slave systems' visuo-spatial sketch pad, episodic buffer and phonological loop, controlled by the central executive.

1) in the stages of processing dimension, perceptual and cognitive tasks use different resources than response selection and execution;
2) spatial activity uses different resources than verbal or linguistic activity;
3) the modalities dimension, different resources are used for auditory and visual perception
4) visual channels are divided on focal and ambient vision
And the main argument of the theory is "to the extent that two tasks use different levels along each of the three dimensions, time-sharing will be better" [43]. The model provides an account on how different elements of the human information processor, like attention, perception, working memory, response selection and execution interact between each other. This theory is also based on evidence from cognitive neuroscience where we can see that different modalities have different locations in the human brain, like primary auditory cortex is involved with auditory perception[] and the visual perception is processed in the occupational lobe[].

However, mental workload can be influenced by the initial perception of the task at hand or the 'appraisal' of it. Similarly to MW appraisal is complex and multidimensional concept[15, 33] that is not well defined.

2.2.2 Measuring Mental Workload

The concept has been explained differently by different authors, and inferences made from various empirical measures which can be divided on primary, secondary, subjective and psychophysiological measures. There are also analytical techniques but we are not concerned with them in this dissertation.

Primary and secondary task mesures

Primary measures rely on operator performance to predict workload. However, a limitation of using primary measures alone is that an operator can spend high amount of effort but this may not be apparent from the performance[**?**]. Consequently, primary task measures should be combined with other workload measures. An example performance measures are task completion time, number of errors, and response time. In our case we will use a combination of all but
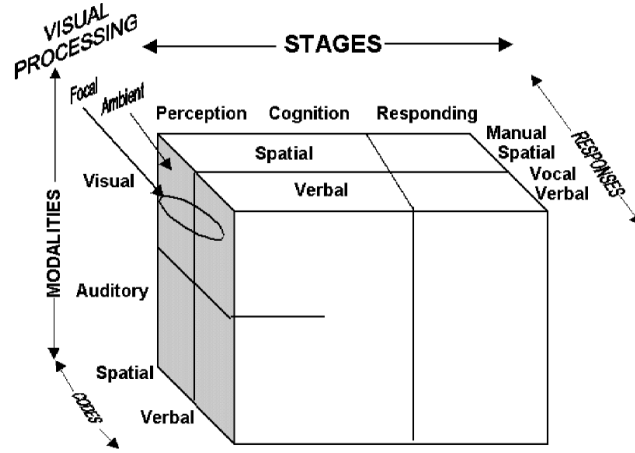
Figure 2.2: Wickens 4-D multiple resources model

secondary task measures. Secondary task involves inclusion of a additional simple task to the primary one, which is done concurrently, if the primary task has low or moderate demand and the level of workload cannot be inferred only from the primary measure. It is used to detect when operators performance deteriorates and this is due to workload overload. However, as we are not going to use secondary measures, more explanation will provided for the other types of measurements.

Subjective measures

Subjective measures use rating scale and are based on operator opinion of their perceived workload during or after completion of task. They are preferred method for WL estimation because they are easy to administer, cheap, and with high face validity. They are classified as uni dimensional and multidimensional. They consist of single subjective scale of workload and multiple scales of types of workload, accordingly. From the unidimensional subjective measures the Cooper-Harper[11] is the most popular among ergonomists and cognitive researchers. However, it is designed for the aircraft domain, and therefore a modified cooper-Harper scale[45] was created for use in other domains. However, as mental workload is influenced by different environmental and personal factors[37], therefore the concept of MW should be considered as multidimensional concept, in order to improve diagnosticity. Accordingly, multidimensional subjective scales should be used to better understand the aspects of MW. The most used scales are NASA-TLX[17], SWAT[35] and Workload profile[41]. NASA-TLX is based on rigorous laboratory research, and it includes 6 scales (mental, physical, temporal demand, experienced effort, frustration and performance) consisting of 20 intervals each and it is relatively easy to administer. In addition, a single measure of workload can be weighted, although it is not necessarily required because there is high correlation between weighted and unweighed results[8]. The SWAT subjective scale is also widely used, however the process of implementation is laborious and more complex than the other

subjective scales. The other popular scale is Workload profile(WP)[41] scale which is basen on the Wickens multiple resources model, and asks questions about each of the four dimensions proposed by the theory. Hence, it is very useful when combined with multiple resource theory interpretation of the results. Finally, Longo et al. [21] compared the three measures mentioned above in a web browsing/searching task, and observed correlations in the results of the three measures claiming that they measure the same concept of mental workload.

Psychophysical measures

Psychophysical measures are used to give objective data about mental workload by not relying on subjective scales or performance measures . They can be obtained by recording cardiac activity, electrodermal activity, eye function or imaging the brain. These techniques detect the change in the arousal from the autonomic nervous system level which can be inferred to as mental workload. However, different psychophysical measures capture different aspects mental workload[9], therefore consideration should be put in choosing the most appropriate measure for the given task.

Mean heart rate(HR) and heart rate variability(HRV) are one of the most used techniques to infer arousal because it is relatively cheap and easy to administer. However, HR not always correlates to subjective measures of MW[16] and because of this HRV can be considered as more valid measure. Moreover, the beat to beat interval of the heart can be measured using different statistical approaches[5], like, standard deviation from heart beat intervals. Measurements of eye activity, like blink rate, pupil diameter are also being found to correlate to MW. Furthermore, increase in pupil diameter is correlated to rise in arousal [19], and Beatty claimed that it has high sensitivity [3] and it can be used to distinguish between data-limited and resource limited processing, which can make it very useful in the HCI field. However, incoming light at the eye can change the pupil diameter, which is a process unrelated to the task, thus influencing the measurements, therefore it is suitable for experiments in controlled environment.

Finally, a number of brain imaging techniques are used to obtain measurements from the brain activity, including electroencephalography(EEG), functional near-infrared spectroscopy(fNIRS), functional magnetic resonance imaging(fMRI), however these will be discussed later in section "Brain sensing".

2.3 Emotion processing

first reptilo plex, than emotion in pfc, then the approaches valence vs sad, frustrated, etc what is emotional valence - emotional valence with arrousal - emotional appraisal studies SAM and other measures

2.4 Brain sensing

Initially, brain sensing was used in for medical purposes, however in recent years it has been used in other fields, like cognitive psychology, and lately, for Brain-

Computer interface(BCI). We review only brain sensing techniques are suitable for measuring complex cognition and are non-invasive so it can be used in HCI experiment. First of all, the functional magnetic resonance imaging (fMRI) is measuring the haemoglobin oxygenation and deoxygenation in the brain which is referred to as blood oxygen level-depended contrast (BOLD). It uses a large magnet which causes a strong magnetic field and a short radio-frequency pulse is emitted, in order to detect areas of activation in the brain. Furthermore, MRI has high spatial resolution which detects activation in brain regions with up to 1mm precision and is suitable for destinguishing particular brain regions of activation for the studied task. Also, its temporal resolution has 2-3 seconds delay, as it takes some time for the neural activity to occur and be detected. However, during experiments participants should stand still and even the slightest movements can cause artefacts and distort the signal. Therefore, this technique is not suitable for HCI evaluations because participants cannot physically move. Another widely used brain imaging technique is the EEG which measures electrical changes at the surface of the scalp. It is non-invasive technique, which makes it suitable for cognitive[] and HCI research[]. Waveforms with different bands are calculated from the electrical signal that can later be analysed. The strength of EEG is its temporal resolution as it can detect changes in brain activity with accuracy of a few milliseconds. However, it is highly susceptible to motion artefacts, and even the movements of the fingers can cause deformation of the signal. Consequently, it is not suitable for web interface evaluations because when users type on the keyboard or use the mouse will cause considerable distortions of the brain signal. A more recent brain imaging technique is the functional near infrared spectroscopy (fNIRS) which unlike EEG is an optical-imaging modality. fNIRS can detect cerebral hemodynamics by calculating the oxygenated haemoglobin(Hbo) and deoxygenated haemoglobin(Hbr). It uses infrared light (usually 650-1000nm wavelenght) which is fired to the participants skull using emitter-detector pairs consisting of infrared LED emitter, and infrared sensors. They usually operate with two or more wavelenghts. Moreover, because Hbo and Hbr have different light absorption coefficients the concentration of Hbo and Hbr can be calculated using modified Beer-Lamberts law[14]. Also, the sensors are strategically placed

Such brain sensing modalities are the fMRI, EEG, fNIRS discuss the different brain imaging techniques, - MRI, EEG, fNIRS. Because fnirs is portable, n

2.4.1 Fnirs and Mental Workload

- PFC is involved in cognitive control(higher cognitive abilities) functions [**?**, **?**, **?**] Highlight studies involving mental workload. More specifically, there is a positive correlation between the increase of oxygenated blood and the increase in cognitive WL- -limitations has low temporal resolution

2.4.2 fNIRS and Emotional Valence

It is still not proven that activation in the left hemisphere is responsible for the processing of positive emotion. gender differences, left-right handedness 1. Emotional induction - when pictures or videos or other method is used to trigger certain kind of emotion Emotional regulation - apprisal - reapprisal - kogato si spomnime i preocenim dadena emociq (cognitively reexamine the meaning of

13

emotional events) - anticipation of expected outcomes - In this study we have combined "hot" emotional control with cold control of attention and memory, as suggested by Kevin N. Ochsner and James Gross [**?**] -Richard Davidson first a comprehensive research in the topic cortical assymetry

2.5 Summary

Because not many researchers[] do not take in consideration the emotional state(positive or negative) which can be related to aproach and avoid motivation during task execution, and how it influences the performance of the operator and perceived workload, we have combined multidimensional subjective scale (NASA-TLX) with emotional valence scale (SAM). By combining the these subjective measures and comparing them the objective measure of mean Hbo from the fNIRS device we expect to gain better understanding of the operator performance during the web form filling task. Arrousal should correlate with skin conductance. We are aiming to reduce to imposed load by the task by minimizing the visual search(condition3) and aiding the episodical memory by placing a description in the beginning of the form.

Chapter 3

User study

3.1 Hypothesis and expectations

Based on the literature review we state the following hypothesis:
1) There is statistical difference between the three web forms
2) Subjective data from NASA-TLX correlates to objective data from the fNIRS
3) The difference in left and right hemisphere activations correlate to subjective SAM scale of emotional valence

3.2 Method

Describe why I have used the following methods, including the perceived benefits of your approach[22].

3.2.1 Participants

3.2.2 Apparatus

Laptop computer

The experiment was executed on 15" laptop, HP probook 450 with screen resolution 1366x768. The participant was presented with a screen with links to the three different videos and web forms. They were instructed by the researcher to manually start certain condition or video.

fNIRS

Picture of fNIRS
730nm and 850nm wavelengths will be collected for each voxel, eliminating the ambient light. . However, if the data acquisition computer does not have enough bandwidth, one or more of the quadrants can be disabled to maintain 2Hz sampling rate.

Empatica

Picture of Empatica E3

Before attaching the Epatica E3 the skin was treated with alcohol for better conductivity.

3.2.3 Materials

NASA TLX

SAM

Web forms

Video capture

3.2.4 Design

The study used repeated measures within subjects designThe three variations of videos and the web forms were counterbalanced, in order to Performance measure - characters written in the description field

3.2.5 Procedure

First, participants were asked to read and sign information sheet and consent forms. Second, the Empatica E3 and fNIRS devices were cleaned then equipped and started. Third, participants were briefed about the procedure of the experiment, and it was explained how to fill the subjective scales. Also, because of ethical considerations that the participant should not enter personal data in the web form, a fake personal credentials were provided, that she should fill in the web forms. Fourth, after the video capture, fnirs and empatica devices started recording data, participants were asked to open one of the three videos, depending on the counterbalancing table. After the video was finished, participants fill SAM subjective scale. Fifth, there was approximately 2 minute waiting period so that participant's memory is not fresh before she was asked to open one of the three web forms, again depending on the order from the counterbalancing table. Finally, after participant has completed the web form, the SAM and then the NASA-TLX scales are given to be completed, accordingly. This process was repeated three times, following the within subjects experimental design. Before each experiment, the clocks between the two computers (the one that was obtaining the information and one that participant worked on) were synchronized. Also, timestamps using the Cobi Studio software manual markers were created in the beginning and end of each condition and video.

3.2.6 Data Analysis

The fNIRS data was analyzed with NIRS-SPM[48]

Signal acquisition

What is the task - it is not mental aritmetic, or mental imagery, it is complex web form filling task.

Preprocessing

Instrument noise was reduced by placing a hat over the fNIRS headband, in order to block external light. First, low-pass filter with cut off frequencies of 0.1 Hz, was used in order to remove physiological noise, like heartbeat and blood flow movement that is not associated with brain activity or Mayer waves. Then, the NIRS signal was processed with modified Beer-Lambert law[12], in order to calculate oxygenated, and deoxygenated hemoglobin values. Finally, to remove motion artefacts, the correlation based signal improvement(CBSI)[13] method was applied to the data.

Feature Extraction/selection

After data preprocessing we should choose what features we will use to evaluate our data - mean values of oxy-Hb, mean values of oxy-Hb for all channels, skewness, variance

3.3 Results

3.3.1 Mental Workload between the three web forms

fNIRS data

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in mean Hbo values between the three web forms. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $X^2(2) = 0.195, p = 0.907$. There was no significant statistical difference in the mean Hbo between the 3 web forms $F(2, 20) = 3.400, p < .054$, partial $\eta^2 = .254$ with mean Hbo decreasing from 0.2377 (SD = 1.19) in index3 to -0.1166 (SD = 0.82) and -0.117 (SD = 1) for index2 and index1 respectfully. No statistical significance was found when comparing the
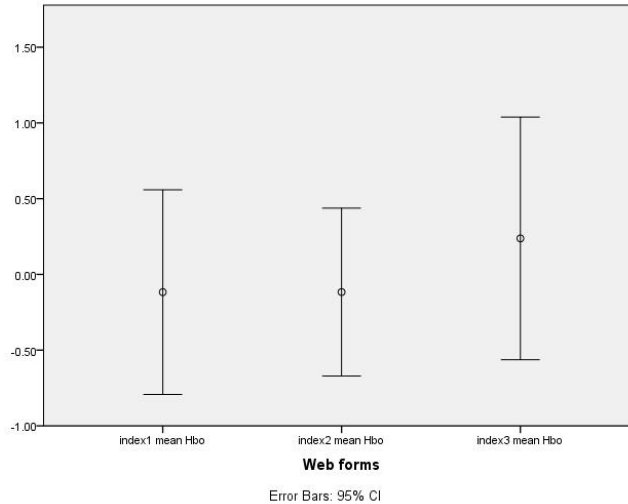

Figure 3.1

17

means of Hbr between the three conditions $F(2, 20) = 2.044, p < .156$, partial $\eta^2 = .170$ where index2 had the highest Hbr mean 0.05 (SD = 0.85), index1 with -0.07 (SD = 0.96) and index3 with the lowest Hbr mean -0.36 (SD = 1.43). Also, a repeated measures ANOVA test was conducted to elicit significant statistical differences between mean Hbt between the three web forms, however no statistical significance was found $F(2, 20) = 0.685, p < .516$, partial $\eta^2 = .064$ where index2 had the highest Hbt mean -0.08 (SD = 0.49), index3 with -0.13 (SD = 0.58) and index1 with the lowest mean Hbt -0.19 (SD = 0.49).

NASA-TLX

There was no statistical significance between each of the NASA-TLX scales, including the total score as assessed by one way repeated measures ANOVA. Perceived mean mental demand was lowest for index1 9.15 (SD = 4.94), index2 had slighly higher mean 9.40 (SD = 4.68) and index3 has the highest scores 10.8 (SD = 5.38). Also, mental demand had a strong positive correlation with Total tlx for the 3 conditions $r(18) = 0.652, p = 0.002$, $r(18) = 0.738, p < 0.001$, and $r(18) = 0.741, p < 0.001$ for index1, index2 and index3 respectfully. The total calculated value for the NASA-TLX was highest for index3 7.07 (SD = 3.22) decreasing to 6.92 (SD = 2.95) for index1 and 6.47 (SD = 3.11) for index2. There was a moderate positive correlation between mental demand scales and task completion times between the three conditions $r(18) = 0.487, p = 0.030, r(18) = 0.484, p = 0.030, r(18) = 0.638, p = 0.002$. Also, mental demand scales had a moderate positive correlation with the total NASA-TLX between the three conditions $r(18) = 0.652, p = 0.002, r(18) = 0.738, p = 0.0005, r(18) = 0.741, p = 0.0005$

Table 3.1: NASA-TLX mean scores

	Index1	Index2	Index3
Mental demand	9.15 (SD = 4.94)	9.40 (SD = 4.68)	10.8 (SD = 5.38)
Physical demand	4.05 (SD = 4.08)	2.90 (SD = 3.21)	3.90 (SD = 3.65)
Temporal demand	7.40 (SD = 4.49)	7.65 (SD = 5.79)	6.55 (SD = 4.91)
Performance	6.65 (SD = 3.79)	5.60 (SD = 3.62)	6.20 (SD = 3.86)
Effort	8.15 (SD = 4.58)	7.35 (SD = 4.68)	8.20 (SD = 5.30)
Frustration	6.10 (SD = 5.11)	6.00 (SD = 3.66)	6.75 (SD = 5.22)
Total	6.92 (SD = 2.95)	6.47 (SD = 3.11)	7.07 (SD = 3.22)

SAM - arousal scale

The perceived arousal was lowest for index1 2.8 (SD = 0.95) incresing to 2.95 (SD = 1.05) for index2 and to 3.15 (SD = 1.18) for index3 respectfully. No statistical significance was found when comparing the means between the three conditions $F(2, 38) = 2.462, p < 0.099$ partial $\eta^2 = .115$ using one way repeated measures ANOVA. However, after running post hoc test without adjustments(LSD) a statistically significant difference was found between index1 and

index3 $p = 0.049$. Also, the time to complete index1 and index2 positively correlated to perceived arrousal for index 1 and index2:$r(18) = 0.551, p = 0.012$ and $r(18) = 0.473, p = 0.035$. However time to complete index3 does not correlate to perceived arrousal of index3 $r(18) = 0.269, p = 0.252$

3.3.2 Emotional Valence

fNIRS differences

WEB FORMS
The fnirs Hbo valence differences was highest for index3 -0.12 (SD = 1.25)
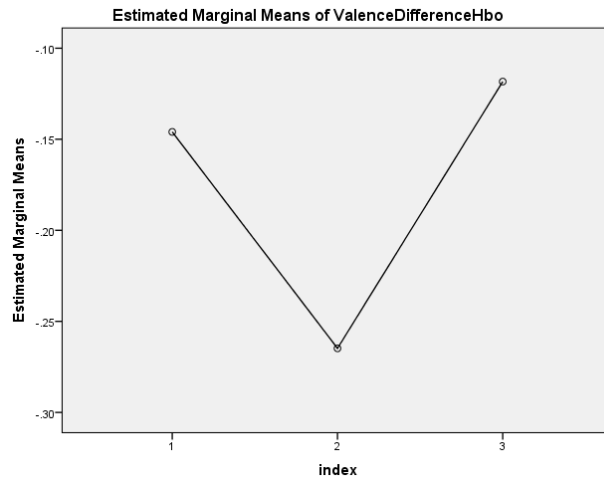

Figure 3.2

decreasing to -0.15 (SD = 1.31) for index1 and the lowest value was for index2 -0.26 (SD = 1.13). The Hbr valence differences were highest for index2 0.34 (SD = 1.18) decreasing to 0.18 (SD = 1.39) for index3 and to 0.10 (SD = 1.68) for index1. The Hbt mean valence difference values for index1 were lower -0.45 (SD = 0.79) compared to index2 0.06 (SD = 0.56) and index3 0.05 (SD = 0.87) respectfully. There was no statistical significance as assessed by one way repeated measures ANOVA between the three conditions for Hbo valence differences: $F(2, 20) = 0.392, p < 0.681$, partial $\eta^2 = .038$, Hbr valence differences: $F(2, 20) = 0.418, p < 0.664$, partial $\eta^2 = .040$ and Hbt valence differences: $F(2, 20) = 0.302, p < 0.743$, partial $\eta^2 = .029$. Also, there was strong positive correlation between temporal NASA-TLX scale of index1 and the Hbo valence differences of index1 $r(9) = 0.766, p = 0.006$, however, there was no correlation found between index2: $r(9) = 0.581, p = 0.061$ and index3: $r(9) = 0.218, p = 0.519$.
VIDEOS
The mean Hbo valence difference for video3 was the highest with 0.17 (SD= 0.25) compared to video1 with -0.01 (SD = 1.32) and video2 with -0.8 (SD = 1.20). In contrast, mean Hbr valence difference values for video3 were the lowest with -0.25 (SD = 0.47) compared to video1 0.30 (SD = 1.20) and video2 0.32 (SD = 0.89). For the mean Hbt valence difference values video1 was the highest

with 0.18 (SD = 0.71) decreasing to 0.07 (SD = 0.64) for video3 and to -0.06 (SD = 0.35) for video2.

A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in Hbo, Hbr and Hbt valence differences between the three videos. There was no significant statistical difference in the mean Hbo valence difference between the 3 videos $F(2, 20) = 0.051, p < 0.951$, partial $\eta^2 = .005$, the mean Hbr valence difference: $F(2, 20) = 0.062, p < 0.940$, partial $\eta^2 = .006$ and the mean Hbt valence difference: $F(2, 20) = 0.522, p < 0.601$, partial $\eta^2 = .050$. There was no correlation found between Hbo valence differences and SAM emotional valence subjective scale for the three videos $r(9) = -0.490, p = 0.126; r(9) = 0.095, p = 0.781; r(9) = 0.496, p = 0.121$.

SAM emotional valence

WEB FORMS

The perceived mean emotional valence for index1 was the lowest with 3.1 (SD = 0.97) increasing to 3.4 (SD = 0.99) for index2 and to 3.7 (SD = 0.98) for index3. A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in SAM emotional valence scale values between the three web forms. The assumption of sphericity was met, as assessed by Mauchly's test of sphericity, $X^2(2) = 0.446, p = 0.800$. There was no significant statistical difference in the SAM emotional valence scale between the 3 web forms $F(2, 38) = 2.803, p < .073$, partial $\eta^2 = .129$ with mean SAM emotional valence increasing from 3.1±0.97 in index1 to 3.4±0.99 and 3.7±0.98 for index2 and index3 respectfully.

VIDEOS

The perceived mean emotional valence for video3 was the highest with 3.1 (SD = 1.07) decreasing to 2.9 (SD = 1.33) for video1, and 2.7 (SD = 0.98) for video3. There was no statistical significance as assessed by one way repeated measures ANOVA between the three videos for SAM emotional valence: $F(2, 38) = 0.792, p < 0.460$, partial $\eta^2 = .040$.

3.3.3   User performance and preferences

The mean time to complete index2 was the lowest 214.88 (SD = 63.81) increasing to 228.79 (SD = 65.19) for index1 and to 231.60 (SD = 83.33) for index3. However, users mostly preferred index3 and index1 with 10 and 9 votes respectively compared to index2 which was preferred by 3 participants. A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in time to complete between the three web forms. There was no significant statistical difference in time to complete between the 3 web forms $F(2, 38) = 0.556, p < .578$, partial $\eta^2 = .028$. Also, the time to complete index2 and index3 had a strong positive correlation with perceived effort(NASA-TLX) for index2 and index3: $r(18) = 0.702, p = 0.016$, and $r(18) = 0.634, p = 0.036,$. However, time to complete index1 does not correlate to perceived effort of index1 $r(18) = 0.216, p = 0.524$

Chapter 4

Discussion

What was the purpose of the study, and then interpretation of the results

4.0.4 Implications for Design

4.0.5 Disadvantages of the study

The study tries to simulate real conditions, and therefore lacks ecological validity because users wait approximately 2 minutes after they have watched the video to start filling the web form. This way they still hold some of the information in their working memory and the study is trying to simulate long term memory recall.

4.0.6 Future work

4.1 Conclusion

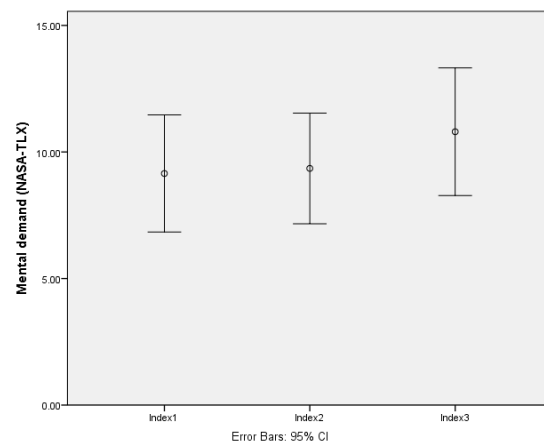In summary, the mental workload is lower for this....

Graphs:

Figure 4.1

References

[1] Alan Baddeley. The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, 4(11):417–423, 2000.

[2] Alan D Baddeley and Graham J Hitch. Working memory. *The psychology of learning and motivation*, 8:47–89, 1974.

[3] Jackson Beatty. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 91(2):276, 1982.

[4] Katie Bessiere, Irina Ceaparu, Jonathan Lazar, John Robinson, and Ben Shneiderman. Social and psychological influences on computer user frustration. *Media access: Social and psychological dimensions of new technology use*, pages 169–192, 2004.

[5] George E Billman. Heart rate variability–a historical perspective. *Frontiers in physiology*, 2, 2011.

[6] D Broadbent. Perception and communications. 1958.

[7] D Broadbent. Decision and stress. 1972.

[8] James C Byers, Alvah C Bittner, Susan G Hill, Allen L Zaklad, and Richard E Christ. Workload assessment of a remotely piloted vehicle (rpv) system. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 32, pages 1145–1149. SAGE Publications, 1988.

[9] Brad Cain. A review of the mental workload literature. Technical report, DTIC Document, 2007.

[10] Ann Chadwick-Dias, Michelle McNulty, and Tom Tullis. Web usability and age: how design changes can improve performance. In *ACM SIGCAPH Computers and the Physically Handicapped*, number 73-74, pages 30–37. ACM, 2003.

[11] George E Cooper and Robert P Harper Jr. The use of pilot rating in the evaluation of aircraft handling qualities. Technical report, DTIC Document, 1969.

[12] M Cope and David T Delpy. System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Medical and Biological Engineering and Computing*, 26(3):289–294, 1988.

[13] Xu Cui, Signe Bray, and Allan L Reiss. Functional near infrared spectroscopy (nirs) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage*, 49(4):3039–3046, 2010.

[14] David T Delpy, Mark Cope, Pieter van der Zee, SR Arridge, Susan Wray, and JS Wyatt. Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in medicine and biology*, 33(12):1433, 1988.

[15] Susan Folkman, Richard S Lazarus, Christine Dunkel-Schetter, Anita DeLongis, and Rand J Gruen. Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. *Journal of personality and social psychology*, 50(5):992, 1986.

[16] Eija Haapalainen, SeungJun Kim, Jodi F Forlizzi, and Anind K Dey. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 301–310. ACM, 2010.

[17] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[18] Caroline Jarrett and Gerry Gaffney. *Forms that work: Designing Web forms for usability*. Morgan Kaufmann, 2009.

[19] Daniel Kahneman. *Attention and effort*. Citeseer, 1973.

[20] Lorna Lines, Oluchi Ikechi, K Hone, and Tony Elliman. Online form design for older adults: Introducing web-automated personalisation. In *Proceedings of HCI, the Web and the Older Population, workshop at HCI 2006*, 2006.

[21] Luca Longo, Fabio Rusconi, Lucia Noce, and Stephen Barrett. The importance of human mental workload in web design. In *WEBIST*, pages 403–409, 2012.

[22] Sheena Luu and Tom Chau. Decoding subjective preference from single-trial near-infrared spectroscopy signals. *Journal of Neural Engineering*, 6(1):016003, 2009.

[23] H Maior, Matthew Pike, Sarah Sharples, and Max L Wilson. Examining the reliability of using fnirs in realistic hci settings for spatial and verbal tasks. *Proceedings of CHI*, 15:3807–3816, 2015.

[24] Horia A Maior, Matthew Pike, Max L Wilson, and Sarah Sharples. Continuous detection of workload overload: An fnirs approach. *Contemporary Ergonomics and Human Factors 2014: Proceedings of the international conference on Ergonomics & Human Factors 2014, Southampton, UK, 7-10 April 2014*, page 450, 2014.

[25] Valerie Mendoza and David G Novick. Usability over time. In *Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information*, pages 151–158. ACM, 2005.

[26] George A Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[27] U Neisser. *Cognitive psychology*. Prentice-hall, 1967.

[28] Jakob Nielsen and Jonathan Levy. Measuring usability: preference vs. performance. *Communications of the ACM*, 37(4):66–75, 1994.

[29] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256. ACM, 1990.

[30] Don Norman. Emotion & design: attractive things work better. *interactions*, 9(4):36–42, 2002.

[31] Donald A Norman and Daniel G Bobrow. On data-limited and resource-limited processes. *Cognitive psychology*, 7(1):44–64, 1975.

[32] Donald A Norman and Stephen W Draper. User centered system design. *Hillsdale, NJ*, 1986.

[33] Edward J Peacock and Paul TP Wong. The stress appraisal measure (sam): A multidimensional approach to cognitive appraisal. *Stress Medicine*, 6(3):227–236, 1990.

[34] Matthew Pike, Horia A Maior, Martin Porcheron, Sarah Sharples, and Max L Wilson. Measuring the effect of think aloud protocols on workload using fnirs. In *ACMCHI*, 2014.

[35] Gary B Reid and Thomas E Nygren. The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology*, 52:185–218, 1988.

[36] Yvonne Rogers, Helen Sharp, Jenny Preece, and Michele Tepper. Interaction design: beyond human-computer interaction. *netWorker: The Craft of Network Computing*, 11(4):34, 2007.

[37] William B Rouse, Sharon L Edwards, and John M Hammer. Modeling the dynamics of mental workload and human performance in complex systems. *IEEE Transactions on Systems, Man, & Cybernetics*, 1993.

[38] Sergio Sayago and Josep Blat. Some aspects of designing accessible online forms for the young elderly. In *WEBIST (2)*, pages 13–17, 2007.

[39] Sergio Sayago, José-María Guijarro, and Josep Blat. Selective attention in web forms: an exploratory case study with older people. *Behaviour & Information Technology*, 31(2):171–184, 2012.

[40] Ben Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*, volume 3. Addison-Wesley Reading, MA, 1992.

[41] Pamela S Tsang and Velma L Velazquez. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3):358–381, 1996.

[42] Christopher D Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002.

[43] Christopher D Wickens. Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):449–455, 2008.

[44] christopher D Wickens and J M Flach. Information processing. *Human factors in aviation*, pages 110–156, 1988.

[45] Walter W Wierwille and John G Casali. A validated rating scale for global mental workload measurement applications. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 27, pages 129–133. Sage Publications, 1983.

[46] Luke Wroblewski. *Web form design: filling in the blanks*. Rosenfeld Media, 2008.

[47] Erik Wstlund, Torsten Norlander, and Trevor Archer. The effect of page layout on mental workload: A dual-task experiment. *Computers in Human Behavior*, 24(3):1229 – 1245, 2008. Instructional Support for Enhancing Students' Information Problem Solving Ability.

[48] Jong Chul Ye, Sungho Tak, Kwang Eun Jang, Jinwook Jung, and Jaeduck Jang. Nirs-spm: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage*, 44(2):428–447, 2009.

Appendix