Quantifying Negative Affect

Usability Testing to Observe the Effect of Negative Emotions on User Productivity Through the Use of Biosignals and OCC Theory

Gloria Washington
School of Computing, Clemson University, Clemson, SC, U.S.A.

Keywords: OCC, Biosignal, Negative Affect.

Abstract: Humans sometimes experience negative emotions caused by electronic devices that impede their task(s). User experience researchers have examined technology-caused negative affect by collecting task performance metrics, user feedback, and/or human physiological data like skin temperature or blood pressure for more insight. Much research has been done to determine the amount of negative affect produced by the humans during these events. However, these methods usually require the user to self-report their negative feelings through Likert scales, pressure-sensitive devices or other manual methods. Task performance measures have also been used in lieu of asking a user what they feel. In this research, we adapt OCC Theory for use with physiological data for quantifying negative affect in human-computer interactions, along with asking a person how they feel about an application. In addition, we observe how negative affect amounts impact task performance measures in a usability study by adding random system delays into an application to induce negative feelings. Results from this work showed productivity does not always degrade when negative feelings are experienced by a user. In addition, some types of negative affect may have the opposite effect and allow a user to increase their performance under the right conditions.

1 INTRODUCTION

On a typical day, humans will come into contact with some type of electronic device more than ten times (Modapt, and Morrissey, 2011). During these interactions, technology causes the user to experience some sort of frustrating event 18% of the time. User experience researchers have studied how to reduce the occurrence of these events through usability testing. During usability testing, users will interact with a device and be asked to perform certain tasks while a usability professional captures data like utterances, human physiological data, or task performance measures (Ward & Marsden, 2003). In addition, users may be asked to fill out a user survey after the test, sometimes in the form of a Likert scale, to determine the amount of negative or positive feelings experienced during the test. Human physiological data is gathered to understand more fully how a person is feeling about a device or application. Productivity metrics also provide insight into what tasks a user can perform quickly or slowly and/or easily or with difficulty within an application. Usability experts use all of this data to find out what functionality works well and what needs refinement.

Negative feelings or affect caused by technology have been studied extensively by researchers in both affective computing and human factors engineering. Theories like Goal theory, Appraisal theory, or Frustration theory have helped early researchers of human-computer interaction to shed light on why these emotions occur in human-computer interactions (Freud, 1922 & Scherer, 2001). What they found through various studies is that negative emotions caused by technology can occur due to a number of factors including how determined a person is at completing a task, how sure a person is of themselves in completing a task, and/or the significance of the event that caused the negativity to occur (Bessiere *et al.*, 2006).

1.1 OCC Theory of Emotions

In 1988, Ortony, Clore, and Collins (OCC) developed a structure for modelling human emotions (Ortony, Clore, Collins, 1988). Unlike other emotional theories created before it, this model was

Washington G..
Quantifying Negative Affect - Usability Testing to Observe the Effect of Negative Emotions on User Productivity Through the Use of Biosignals and OCC Theory.
DOI: 10.5220/0005246200830089
In *Proceedings of the 2nd International Conference on Physiological Computing Systems* (PhyCS-2015), pages 83-89
ISBN: 978-989-758-085-7

designed specifically for characterizing emotions by situations that cause them to occur. Also, emotions do not occur until an individual has reached and surpassed a unique threshold inherently set within a person. Essentially people's perception of events causes them to experience situations differently. For an example, a person that is confined to a time-constraint is more likely to experience a greater amount of frustration if they are impeded from completing a task, whereas someone without a time limit may experience less.

In OCC theory, emotions are grouped into 22 groups are called "emotion types". Each emotion type has different factors affecting the intensity of that type. Factors that tend to increase emotion intensity often increase the potential for other emotions to occur in that emotion type. For instance, a person may have a deadline to submit his/her tax return by 12:00AM on April 15the and may experience a blackout that causes his/her Internet to be down, thereby causing the person to come dangerously close to incurring a penalty because the return is late. In this situation, the person may initially feel extreme anger towards his/her self or even "Mother Nature". However, as the consequences of the event unfold, he/she may begin to feel fear over missing the tax submission deadline.

Also in OCC theory, people perceive the world to be events, agents, and objects. Emotions occur due to consequences of events, actions of agents, or aspects of objects. In the example given about the person and his/her tax return, the Internet/weather is the agent, the action of the agent is the blackout, and the late penalty from the IRS is the consequence of the power outage.

Humans often have competing and conflicting goals that may impact the intensity of an emotion. Unfortunately, OCC theory does not assess the consequences of multiple, competing goals. However, it does address how to determine emotion intensity for each goal expressed by a human.

In OCC Theory, the intensity of emotion experienced by an individual pertains to: the congruence of an event's consequences with one's goals (i.e. the user is pleased when his computer "helps" him by automatically typing a word into a report, but displeased if it inserts an incorrect word); the consequences of actions of agents(one's self, people, or inanimate objects such as computers) according to some standard (i.e. a person is displeased when he realizes he has lost his report due to his failure in saving the document); and the consequences of people's attitudes or disposition to like or dislike certain objects or aspects of objects (i.e. people's attitudes about root canals causes the idea of going to the dentist to be unappealing.) Equation (1) shows the original OCC structure for determining emotion intensity. This method, described in the next section, is modified for determining frustration intensities using bio-signal data.

1.2 Contribution to Physiological Computing

There has been much research on the use of physiological data in usability studies to understand negative affect (Westerman *et al.*, 2006). This work does not purport that gathering and analyzing physiogical data in usability testing is new. However, quantifying negative affect from physiological data using OCC Theory is a novel approach. Additionally, this work does not suggest OCC theory is the best psychological theory for describing negative affect that occurs in human-computer interactions. However, this work examines OCC theory because it is a psychological theory that includes a computational model for quantifying negative affect that may occur in human-computer interaction. This work builds upon research related to negative emotions caused by technology, OCC Theory and negative emotion modeling, and explores two main themes:

- Determining the amount of negative emotion experienced by a user in a usability test through the use of the OCC Theory of Emotions
- Determining the unique effect of negative emotion amounts on user productivity.

2 RELATED STUDIES

Negative emotions caused by electronic devices include frustration, annoyance, anger, and/or stress. Frustration, as described by Freud is any event that occurs and impedes a user from completing a task (Freud, 1922). Bessiere studied user frustration (defined as frustration during computing) in the workplace by having participants keep diaries and log their experiences as they interacted with a tool (Bessiere *et al.*, 2006). In addition, subjects filled out Likert scale surveys to report the frustration they experienced after the study. From this work, the Computer Frustration Model was developed to help understand the relationship between problems encountered by workplace computer users and the frustration and mood of the users. Strong predictors of negative mood were strongly linked to a person's self-efficacy or belief they can accomplish the task on the computer, the severity of an interruption impeding a person from completing a task, and the importance of the goal to the person.

System delays are found to be the most common task inhibitor and computer users seem to exhibit the most negative feelings when they occur (Scheirer, 2001). In 2004, affective computing researchers Picard and Klein used system delays to study the physiological effects of stress/frustration on the human body (Picard & Klein, 2005). In their research they captured blood pressure and heart rate data along with the use of a hidden Markov model (HMM) classifier (Ghahramani, 2001) to allow the computer to respond to negative affect exhibited by participants in the study. In addition, other researchers have explored using human physiological data to better understand negative affect in human computer interactions (Klein, 2001; Hazlett, 2003; Riseberg, 1998; Picard 1997 & 2003, Scheirer, 2001).

In most usability studies frustration is self-reported; however researchers have begun to explore computer hardware that is able to capture the stress experienced by a user through pressure sensitive mice and keyboards (Yuan & Picard, 2013 and Hernandez *et al*., 2014). Rajendran (Rajendran, 2011) calculated frustration-index scores generated from student log data and time information gathered from various activities. These frustration index scores were verified against frustration amounts self-reported by students via a pop-up window in an intelligent tutoring application.

The previous studies mentioned rely on psychological theories like Frustration theory, goal theory, and appraisal theory to understand the amount of negative emotions that occur when a computer unexpectedly blocks a user from completing a task. This work, however, uses OCC Theory of Emotions (called OCC) to understand the amount of negative feelings produced by person being blocked from completing a task or goal (Ortony, Clore, & Collins, 1988). OCC says that negative compound emotions and attribution emotions occur as a result of consequences of an action attributed to an agent. In this work, the agent is the computer and the action is the task-inhibitor or blocker.

In 1993, Elliot (Elliott, 1992) implemented an artificial intelligence application called TaxiWorld that utilized an emotional model called the Affective Reasoner based on OCC theory. In this application, users would navigate their taxi through a world based off of the Chicago area and experience various emotions including anger. The other taxis in the program would react using the underlying emotional model as various situations presented itself to a user and his/her taxi.

Katsionis and Virvou (Katsionis & Virvou, 2005), used OCC theory to create an instructional technology tool for teaching English to Spanish speaking students. In this tool, the emotional model would learn from student input and areas within the instruction where students needed more help. The emotional model used by Katsionis and Virvou calculated intensities for performance metrics related to English translations provided by the student.

The previous studies described use system delays, human physiological signals, task performance measures, various psychological theories for understanding negative affect, and/or user feedback data to study user productivity and/or negative affect in human computer interactions. This research uses the OCC, along with human physiological indicators of negative affect to determine the amount of affect experienced by users in a usability study. In addition, this study calculates user performance metrics and determines what amounts of negative affect degrade task performance.

3 METHOD

To study the relationship between the amount of negative affect experienced by a user and task performance, an experiment was performed to gather human biological data, productivity metrics, and user feedback. In addition, the original OCC model was adapted for calculating amounts of negative emotion experienced by users.

3.1 OCC Adaptation

The original OCC computational model that is included in the theory is shown in Equation 1. This model says that an emotion has not occurred unless it has surpassed a person's unique internal threshold. Therefore, intensities of an emotion can be calculated once it has surpassed a person's unique threshold.

To adapt this computational model for human biological data it is necessary to come up with a person's unique threshold. Upper and Lower, as shown in Equation 2, do just that. Upper and Lower measures account for a user's normal physiological

if (emotion-potential) > (emotion-threshold)
then
(emotion-intensity) =
(emotion-potential) – (emotion-threshold) (1)
else
(emotion-intensity) = 0;

behaviour while they are interacting with an electronic device. An example of this is if a usability participant is frustrated about a getting an expensive parking ticket before starting a testing session. Her/his physiological signals may include outliers due to increased heart rate, skin temperature, blood pressure that often accompany negative feelings due to anger (Hazlett, 2003). The values of the Upper and Lower would help to find a user's internal threshold for a negative emotion to occur during the usability test not any other negative emotions that may have occurred prior to testing.

Upper = Mean + Standard Deviation
Lower = Mean – Standard Deviation (2)

if (bio-signal > Upper)
then
intensity = bio-signal – Upper
else if (bio-signal<Lower)
then (3)
intensity = Lower-bio-signal
else
intensity = 0

Intensity values and their interpretations are shown in Table 1. For simplicity reasons in the table, the term frustration is used to encompass all negative emotions experienced by a subject in this usability test. We understand that frustration has a specific definition that is related to a goal-blocking event.

Table 1: Intensity values and their interpretations.

Intensity	Description
0	indicates no frustration has occurred and the user has not surpassed their normal physiological range
1	indicates user has surpassed their threshold and a minimal amount of frustration has occurred
2	indicates a low amount of frustration has occurred
3	indicates a medium amount of frustration has occurred
4	indicates a medium-high amount of frustration has occurred
5	indicates a high amount of frustration

3.2 Experiment

Forty-two participants were asked to interact with a modified online word processing tool, called tinyMCE shown in Figure 1, that included random system delays between mouse and keyboard output. Users were asked to perform a simple word-processing task by creating a flier for the grand-opening of a Coffee shop in the area. The Coffee shop flier had to contain various formats, images, and other information about the opening. The total tasks to complete by the user were 25. At any time during the study, users were given the option to skip a task and go on to the next task if they did not want to complete it. Users were asked to wear the Bio-Pac harness device around their chest that gathered heart-rate, skin temperature, posture, and various breathing metrics every 4 milliseconds. In addition to this, users were asked to be as vocal as possible when working in the tool and fill out a post-study survey that included a Likert scale to indicate their overall experience with the tool.

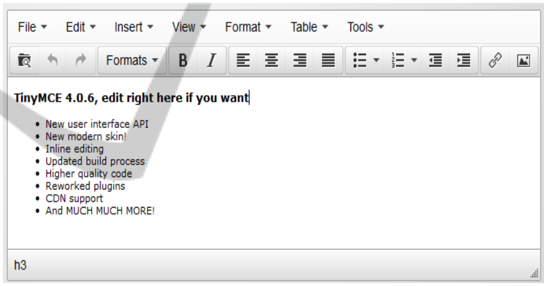

Figure 1: interface of modified word-processing tool.

Metrics captured during the study included:

- number of times a user skips a task
- number of typos
- consecutive number of typos
- number of formatting errors
- number of uncompleted tasks
- number of times student did not follow directions
- time to complete study
- number of skipped tasks
- total number of intensities #1s, 2s, …, n where intensities = [1 to n]
- total number of intensities that decreased task performance
- total number of intensities that increased task performance
- overall intensity for a 4ms
- overall intensity for a session
- number of intensities task performance unchanged

Equation 2 & 3 was used to calculate the intensities for the biosignals gathered by BioPac: heart rate, skin temperature, blood pressure, and breathing metrics. In order to combine these signals, a normalization process was used to convert the signals to the same scale. This transformation insured that the intensities were within the same range, between 0 and 5. Zero through five was chosen because typical usability scale surveys contain a Likert scale from 0 to five for participants to label the amount of negative affect experienced during a test.

4 RESULTS

Initially 44 subjects participated in the study; however two of the individuals were excluded because task performance data was missing or incomplete. Therefore, there were 20 females and 22 males, aged 18-45. Average time to complete the study was about 10 minutes. Only one participant opted to skip a task and stick with the decision. 88% of users experienced a higher "number of typos" and "consecutive typos" than any other measure during the study. The measures with the lowest numbers were "number of skipped tasks" and "uncompleted tasks". These measures indicate users had the most trouble with typing in the interface. However, users did not seem to have trouble with formatting text or adding images in the modified tool.

A majority of the users, 98%, experienced frustration intensities between 1-3. Only one person experienced an intensity of zero indicating no frustration. Also, one participant in the study experienced a frustration intensity of 4. He/she was the oldest participant in the study at 45 years of age. 15%, of the subjects in the study experienced an intensity, either from 1-3 that caused their task performance to decrease. Interesting to note, 55% of the users experienced an intensity that caused task performance to remain unchanged; meaning it did not decrease nor increase. Furthermore, 30% of the users of the modified tool experienced an increase in task performance.

Looking deeper at the category of subjects that experienced unchanged task performance, 20 of the 42 participants experienced a frustration amount of 3 indicating a medium amount of frustration. Whereas, only 13 of the 42 subjects experienced frustration levels from 0-2; indicating minimal to no frustration. Furthermore upon examining the category of subjects that experienced increased productivity, 19 of the 42 subjects experienced medium frustration and 23 of subjects experienced minimal to no frustration. Lastly, analyzing the category of subjects that experienced decreased performance found that 31 of the 42 participants experienced a low to medium amount of frustration (a calculated intensity of 2 or 3).

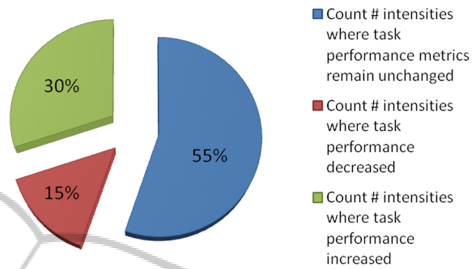

Figure 2: Count of intensities that caused productivity to remain unchanged, decrease, or increase.

To test how well OCC's computational model was at determining a user's overall experience with a tool, we compared the overall session intensity with the user feedback Likert-scale data. 85% of the user-supplied Likert scale data agreed with the overall session intensity calculated by our adapted model.

5 DISCUSSION

Users in this study had more contact with electronic devices due to the age of the population sampled. The average age of the majority of participants was 30 years old. However, more than 50% of the subjects studied were in their early twenties. Many of them have had mobile devices since they were teenagers. Their reactions to task-blockers are somewhat different than an older population that has not been accustomed to electronic devices most of their lives. Furthermore, some of the subjects adopted a competitive stance when it came to system delays. Some of the comments from the participants included "It's easy. I'm used to the delays. I slowed my work down to match the computer" and "I didn't let the delays bother me. I focused my time typing rather than the output on the screen". In addition, some mentioned "the task was easy enough to complete, so I didn't let the problem bother me too much". Perhaps this phenomenon of wanting to beat the system was expressed through their physiological data and the intensities calculated by the modified OCC model. One would assume that higher intensities would result in decreased task performance; however this was not the case in this study. In fact, some users were able to experience "low to medium frustration" without it negatively affecting their performance. Perhaps this explains the amount of subjects that experienced an increase in task performance.

More research is needed to explore the conditions necessary for a person to experience technology- caused frustration or stress without it negatively impacting productivity. Perhaps this information could lead to adaptive interface techniques that optimize a user's productive time based off of intensities calculated from OCC.

The modified OCC computational model in this study uses an upper and lower bound to account for a user's normal behaviour while interacting with an electronic device. However, this upper and lower bound could be found through machine learning techniques that account for outliers in physiological signals caused by events external to a usability test.

6 FUTURE WORK

The study described in this paper will be run again and combined with eye tracking data and mouse pointer data to determine the widgets or functionality that a person is interacting with in a tool. We hope that this will identify aspects of the user interface that need more refinement. Furthermore, we hope that it yields more information for user experience experts to draw from in analyzing the results of a usability study. Along with this, we will test the system with a wider population with various age ranges. We hope that it will help us discover differences in the way older and younger individuals exhibit negative emotions caused by technology and the conditions necessary for increasing productivity in these populations.

Further in the future we will combine Hidden Markov Models to determine the Upper and Lower bound for the modified OCC computational model.

7 CONCLUSIONS

User experience researchers gather various kinds of data including human physiological signals, task performance metrics, and user feedback during/after usability studies. This information helps usability researchers improve the design of a tool by understanding the various causes of technology-induced negative emotions and the events that cause a decrease in user productivity. In this study we wanted to further examine the relationship between task performance and negative emotions caused by task-blocking system delays. We modified the original OCC theory to include an upper and lower bound for calculating the amount or intensity of a negative emotion experienced by a user. We examined how each calculated amount improves, degrades, or does not affect productivity metrics. The usability test from this work showed that users can experience some amount of negative emotion and not have it decrease their task performance. From this study, we determined that more work needs to be done to optimize the time a user is productive, even if they are experiencing some level of negative emotion. Lastly, we believe intensities of negative emotions could give usability engineers extra data to analyze when refining interfaces and/or applications.

ACKNOWLEDGEMENTS

This work was funded through a grant from the Booz Allen Hamilton Center of Excellence. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of Booz Allen Hamilton, Incorporated.

REFERENCES

Bessiere, K. A Model for Computer Frustration: The Role of Instrumental and Dispositional Factors on Incident, Session, and Post-Session Frustration and Mood. Computer in Human Behavior, 2006, 941-961.

Bessiere, K ., Ceaparu, I., Lazar, J., Robinson, J., and Shneiderman, B. Understanding Computer User Frustration: Measuring and Modeling the Disruption from Poor Designs. *HCIL Lab Day 2002*, Technical Report No. 4409, HCIL-2002-18 2002.

BioPac. BioHarness Data Logger Telemetry System AcqKnowledge. 2010 http://www.biopac.com/bioharness-data-logger-telemetry-system-acqknowledge.

Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., and Shneiderman, B. Determining Causes and Severity of End-User Frustration. *International Journal on HCI*, 2004.

Dicks, S. R. Mis-Usability: On the Uses and Misuses of Usability Testing" International Conference on Computer Documentation. 2002, 26-30.

Elliott, Clark. Using the Affective Reasoner to Support Social Simulations. International Joint Conferences on Artificial Intelligence. 1993.

Freud, S. *Beyond the Pleasure Principle*. London: Hogarth Press, 1922.

Ghahramani, Z. An Introduction to Hidden Markov Models and Bayesian Networks. International Journel of Pattern Recognition and Artificial Intelligence. 15(1), 9-42. 2001.

Hazlett, R. Measurement of User Frustration: A Biological Approach. *CHI2003*, Conference on Human Factors in Computing Systems. 2003. 734-735.

Hernandez, J., Paredes, P., Roseway, A. and Czerwinski, M. Under Pressure: Sensing Stress of Computer Users. *CHI2014, International Conference on Human Factors in Computing Systems.* 2014.

Katsionis, G. and Virvou, M. Adapting OCC Theory for Affect Perception in Educational Software. *HCI2005,* 2005.

Klein, J., Moon, Y., and Picard, R.W. This Computer Responds to User Frustration: Theory, Design, and Results. *Interacting with Computers 14,* 2 (2001), 119-140.

Modapt, Inc. and Morrissey & Co. Modapt, Inc./Morrissey & Company Mobile Survey, http://www.modapt.com/wp-content/pdfs/Survey_Results_11Aug2011.pdf.

Moxiecode Systems AB. TinyMCE. 2011. http://tinymce.moxiecode.com.

Ortony, A., Clore G.L., and Collins A.. *A Cognitive Structure of Emotions*. Cambridge University Press, 1988.

Picard, R.W. *Affective Computing*. Cambridge: MIT Press, 1997.

Picard, R. and Daily, S. Evaluating Affective Interactions: Alternatives to Asking What the Users Feel. *CHI2005*, CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches, 2005.

Picard, R. W. and Klein, J. Computers that Recognize and Respond to User Emotion: Theoretical and Practical Implications. MIT Media Lab Tech Report: 538. 2003.

Rajendran, R. Automatic Identification of Affective States Using Student Log Data. *AIED'11*, Proceedings of the 15th International Conference on Artificial Intelligence in Education, 2011, 612-615.

Riseberg, J. Frustrating the User on Purpose: Using Biosignals in a Pilot Study to Detect the User's Emotional State. *CHI2008*, Conference on Human Factors in Computing Systems. 1998.

Scheirer, J. Frustrating the User on Purpose: A Step Toward Building an Affective Computer.2001.

Sherer, K.R., Schorr, A. Ed. Appraisal Processes in Emotion: Theory, Methods, Research. Oxford University Press, 2001.

Ward, Robert D., and Philip H. Marsden. "Physiological responses to different web page designs." International Journal of Human-Computer Studies 59.1 (2003): 199-212.

Westerman, S.J., Gardner, P.H., Sutherland, E.J. Taxonomy of Affective Systems Usability Testing Report. Institute of Physiological Sciences, University of Leeds. 2006.

Yuan, Qi & Picard, R.W. Context-sensitive Bayesian Classifiers and Application to Mouse Pressure Pattern Classification. MIT Media Lab Tech Report, TR-553. 2013.