[Page de garde]

HEC MONTRÉAL

Intégration des mesures neurophysiologiques dans la recherche en expérience utilisateur

par

Vanessa Georges

Sciences de la gestion
(Option Marketing)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

Octobre 2017

Résumé

De plus en plus de recherches en marketing et en gestion tentent d'évaluer de manière quantitative l'expérience utilisateur afin de comprendre les comportements des consommateurs, et ainsi créer l'expérience la plus simple, intuitive et cohérente possible. L'analyse des réactions émotionnelles et cognitives des utilisateurs lors d'une interaction est récemment devenue un avantage stratégique important, passant de l'utilisabilité à l'expérience utilisateur. Cependant, bien que ce changement théorique soit maintenant largement accepté dans la communauté universitaire de l'interaction humain-machine (IHM), l'industrie tarde encore à se l'approprié, notamment en ce qui concerne les méthodes d'analyse et d'évaluation de l'expérience utilisateur.

Ce mémoire par article présente un nouvel outil d'évaluation de l'expérience utilisateur qui vise à contextualiser les signaux physiologiques et comportementaux des utilisateurs, lors d'une interaction humain-ordinateur. Deux études ont été menées en laboratoire afin de développer et valider l'outil proposé. Les contributions à la recherche et les implications pratiques sont discutées.

Mots clés : Expérience utilisateur; oculométrie; cartes de chaleur; design d'interface; signaux neurophysiologiques; charge cognitive; complexité visuelle.

Abstract

More and more marketing and management research are attempting to quantitatively evaluate customer experience in order to understand user behavior, and thus create the simplest, most intuitive, and coherent experience possible. The analysis of the emotional and cognitive responses of users during an interaction has recently become an important strategic advantage, moving away from usability and towards user experience. However, although this theoretical change is now widely accepted in the computer-human interaction community, the implications of this permutation have not yet been reflected in the methods of analysis and evaluation of user experience.

This thesis by articles presents a new user experience evaluation tool that aims to contextualize the physiological and behavioral signals of users during human-computer interactions. Two separate studies were conducted in laboratory to develop and validate the proposed tool. Research contributions and practical implications are discussed.

Keywords: User experience; interface design; eyetracking; heatmaps; neurophysiological signals; cognitive load; visual complexity.

Table des matières

Liste des figures

Liste des tableaux

Liste des abréviations

EDA : Activité électrodermale

CL : Charge cognitive

ECG : Électrocardiographie

EEG : Électroencéphalographie

EMG : Électromyographie

HCI : Human-Computer Interactions

HR : Fréquence cardiaque

HRV : Variation de la fréquence cardiaque

IHM : Interactions humain-machine

IS : Systèmes d'information

ISO : Organisation internationale de normalisation

RSP : Respiration

UX : Expérience utilisateur

Avant-propos

Le mémoire suivant est présenté sous forme d'un mémoire par articles, avec l'accord de la direction du programme de maîtrise ès sciences en gestion de HEC Montréal. Un article ayant fait l'objet d'une publication arbitrée et détaillant une des trois phases du projet de recherche sera présenté dans chacun des trois chapitres du mémoire : l'idéation, la conception et l'évaluation de l'outil.

Le consentement des coauteurs de chaque article a été obtenu afin de présenter ceux-ci dans le contexte de ce mémoire. De plus, le Comité d'Éthique en Recherche de HEC Montréal a approuvé les projets de recherche qui ont servi à produire ces articles.

« Tu nous entends l’Univers? Tu nous entends?
Si tu nous entends, attends-nous! On arrive. »

Remerciements

Tout d’abord, j’aimerais remercier mes directeurs de mémoire. Sylvain Sénécal, pour sa compréhension et ses nombreux conseils tout au long du projet de recherche ; ainsi que pour l’opportunité de rédiger mon mémoire sur un sujet qui me passionne vraiment. François Courtemanche, pour son éthique de travail et pour avoir cru en moi malgré mes propres doutes. Merci d’avoir été un mentor patient, un directeur dévoué, un ami sincère et passionné.

Mes remerciements les plus chaleureux vont à toute l’équipe du Tech3Lab, mes collègues et amis, sans qui ce projet n’aurait jamais pu être réalisé. Merci pour votre aide et votre soutien tout au long de ces trois (parfois longues) années. Merci surtout à Marie-Christine, Shirley, David, Beverly, Élise, Marc, Laurence et Félix.

Introduction

1. Contexte de l'étude

L'évaluation des réactions émotionnelles engendrées lors d'une interaction humain-machine est maintenant considérée comme essentielle pour le développement de produits, systèmes et services qui répondent aux besoins et aux attentes des consommateurs. [1] La norme ISO 9241[1] définit l'utilisabilité comme étant « la mesure dans laquelle un produit peut être utilisé par des utilisateurs précis pour atteindre des buts précis avec efficacité, efficience et satisfaction dans un contexte d'utilisation déterminé ». La notion d'expérience utilisateur fait son apparition dès 1987, alors que Whiteside et Wixon [2] soutiennent que l'utilisabilité ne devrait plus se contenter de mesures de productivité et aussi tenir compte de l'expérience vécue par l'utilisateur.

1.1. Définition de l'expérience utilisateur

Selon l'Organisation internationale de normalisation (ISO-9241-21) l'expérience utilisateur désigne « les perceptions et les réponses d'une personne résultant de l'utilisation ou de l'utilisation anticipée d'un produit, d'un système ou d'un service ». [1] Cette branche de l'interaction humain-machines (IHM) s'inscrit dans une approche multidisciplinaire, reliant l'utilisateur, les machines et les informations contextuelles de cette interaction en un seul domaine d'étude. Selon Hassenzahl & Tractinsky [1], l'expérience utilisateur résulte donc de l'état interne d'un utilisateur (prédispositions, attentes, besoins, motivation, humeur, etc.), les caractéristiques du système conçu (par exemple, complexité, but, facilité d'utilisation, fonctionnalité, etc.) et le contexte au sein duquel se produit l'interaction. S'appuyant sur différents domaines (ex. la psychologie, les neurosciences, l'ergonomie, la biologie, etc.), ainsi que l'informatique et le design industriel, l'expérience utilisateur se définit en trois dimensions : la dimension pragmatique, la dimension hédonique et la dimension émotionnelle. L'expérience

[1] http://www.iso.org

utilisateur suppose donc que les éléments utilitaires, hédoniques et émotionnels de l'interaction soient interdépendants dans la création d'expérience [3].

1.1.1. Dimension pragmatique

La dimension pragmatique de l'expérience utilisateur met l'accent sur les aspects utilitaires de l'utilisation de la technologie, par exemple la facilité d'utilisation et fiabilité du système [1]; puisque ceux-ci auront aussi un impact sur l'expérience globale des utilisateurs.

1.1.2. Dimension hédonique

La dimension hédonique de l'expérience utilisateur considère les aspects hédoniques d'une interaction humain-machines, comme la stimulation (c.-à-d. la croissance personnelle, l'augmentation des connaissances et les compétences), l'identification (c.-à-d. expression de soi, interaction significative avec d'autres) et l'évocation (c.-à-d. l'entretien de soi, la mémoire) [4].

1.1.3. Dimension émotionnelle

Cette dimension s'attarde au rôle de l'affect en temps qu'antécédent, conséquence et médiateur de l'utilisation de la technologie [1]. À noter que l'expérience utilisateur est souvent axée sur la prévention d'émotion négative, par exemple la frustration, la colère et l'insatisfaction plutôt que la création d'émotions positives (c.-à-d. la joie, l'amusement).

1.2. Expérience utilisateur appliquée au marketing

Le marketing est une des fonctions organisationnelles ayant connu le plus de bouleversements au cours des dernières années. L'abondance des technologies de l'information et des médias numériques a profondément modifié les comportements des consommateurs. Ces derniers sont maintenant plus informés et plus exigeants et ont des attentes précises quant aux produits et aux services qu'ils souhaitent obtenir.

Le commerce électronique désigne les activités d'échanges et de transactions effectuées par l'intermédiaire d'Internet, que ce soit pour effectuer un achat, trouver une réponse à une problématique via un moteur de recherche, réserver un hôtel ou partager du contenu entre amis [5]. La qualité de l'expérience utilisateur est un facteur déterminant dans l'exécution de toutes ces tâches, pouvant conduire à un meilleur taux de conversion, d'adoption et de réutilisation de produits et services [6].

Dans un contexte de compétition grandissante, la croissance de l'entreprise passe par la capacité de celle-ci à établir une relation étroite, durable et rentable avec sa clientèle [7]. En d'autres termes, une meilleure identification des besoins et le développement de produits et services détenant une proposition de valeur qui répond plus adéquatement à ces besoins faciliteront l'acquisition, la satisfaction et la fidélisation de clients. Plus la solution proposée répond précisément aux besoins des consommateurs, plus leur satisfaction augmente. Cela implique donc différentes techniques de recherche marketing, passant des méthodes traditionnelles (ex. sondages, entrevues, questionnaires) aux nouvelles approches exploitant la puissance des données quantitatives (ex. les mesures de l'audience d'un site Web, les mesures neurophysiologiques et comportementales).

2. Méthode d'évaluation de l'expérience utilisateur

Le passage de l'utilisabilité à l'expérience utilisateur, un concept plus riche, global et directement lié au but visé par les utilisateurs eux-mêmes [8, 9], requiert l'élaboration de nouvelles méthodologies de recherche et méthodes d'évaluation, permettant aux experts de mieux comprendre et mesurer les comportements et motivations des utilisateurs.

Au cours des dernières années, plusieurs chercheurs, dont Roto et al [10], ont tenté de recenser et classifier les différentes méthodes d'évaluation de l'expérience utilisateur (UX) axées sur la mesure des émotions, afin d'aider les praticiens et chercheurs à sélectionner la bonne méthode d'évaluation, selon les besoins d'évaluation de l'expérience de l'utilisateur; par exemple, le type d'information désiré (données quantitatives ou qualitatives), la phase de développement (formative ou sommative) la période d'étude (court, moyen ou long terme) et le contexte (in situ ou en laboratoire)[2].

[2] http://www.allaboutux.org/

Certaines méthodes sont privilégiées dans certaines sphères du domaine, puisque les besoins d'évaluation de l'expérience utilisateur diffèrent selon différents milieux, par exemple, entre l'industrie et l'académie. En industrie, ces besoins reposent sur l'analyse et la communication adéquate de résultats aux gestionnaires, afin d'améliorer l'expérience des utilisateurs. Les activités et intérêts de recherche universitaire sont centrés sur la validation et la compréhension de phénomènes sur la base d'hypothèses [10].

Ici, nous parlerons de deux types de méthodes d'évaluation de l'expérience utilisateur : les méthodes dites traditionnelles et les méthodes d'évaluation neurophysiologiques.

2.1. Les méthodes traditionnelles

Les méthodes d'évaluation traditionnelles sont nombreuses. Les entrevues permettent aux praticiens et chercheurs de collecter des données qualitatives de tous genres lors d'entretiens avec un utilisateur cible [11]. Les groupes de discussion, ou *focus groups*, désignent des entretiens réalisés avec l'aide d'un modérateur et plusieurs utilisateurs cibles.

Les données résultant de l'observation d'utilisateurs cibles diffèrent grandement selon le contexte de l'étude, ex. l'observation dans l'environnement naturel du sujet ou en laboratoire, ainsi que la technique d'observation utilisée; l'observateur pouvant avoir un rôle plus ou moins actif [12]. Les émotions des utilisateurs peuvent être déduites par l'observateur en analysant le langage corporel, le comportement et les expressions faciales des sujets en temps réel, sans précision absolue.

Les questionnaires, qui peuvent être administrés en ligne ou en personne, servent à recueillir de l'information contextuelle, pragmatique et émotionnelle sur l'expérience vécue des utilisateurs durant leur interaction [11].

Dans tous ces cas, le choix de la population à cibler est primordial afin d'assurer la qualité des données collectées. [11].

2.1.1. Avantages

Les méthodes qualitatives, dues à la richesse des données qu'ils permettent de recueillir, peuvent fournir aux praticiens et chercheurs des réponses au « pourquoi ». Les méthodes traditionnelles d'évaluation de l'expérience utilisateur sont aussi flexibles et faciles à utiliser [11]. Requérant peu de mise en place et d'équipement, ces méthodes peuvent être utilisées dans d'innombrables contextes et à différentes phases du processus de développement. De plus, ces méthodes, qui peuvent être utilisées de manière sommative (une fois l'interface terminée) ou formative (en cours de développement) servent également à générer des idées [11], facilitant ainsi l'exploration dans les premières phases de conception d'une interaction humain-machine. Ainsi, l'utilisateur peut non seulement guider le professionnel dans la conception et l'élaboration d'un système, produit ou service de qualité supérieure, mais peut aussi participer de manière plus active au développement de ceux-ci.

2.1.2. Lacunes

Bien que ces méthodes traditionnelles aient fait leur preuve depuis longtemps et sont utiles aux praticiens et chercheurs en UX, de nouvelles approches s'offrent à nous grâce aux nombreux avancements technologiques. Ces nouvelles approches adressent en particulier les biais méthodologiques et le manque de précision temporelle des méthodes traditionnelles. Plusieurs méthodes d'évaluation traditionnelles, par exemple les questionnaires et entretiens, reposent sur des données autodéclarées afin d'évaluer les états affectifs (p. ex., tristesse, joie, surprise, etc.) et cognitifs (p. ex., charge cognitive, stress, etc.) des utilisateurs, qui sont souvent exposés à des biais méthodologiques, tels que la désirabilité sociale [13], influençant ainsi les performances et comportements des utilisateurs. Plus encore, ces méthodes évaluent la perception des usagers soit pendant ou après l'interaction, ce qui peut soit perturber l'expérience de l'utilisateur ou induire des biais rétrospectifs [14, 15].

2.2. Les mesures neurophysiologiques et comportementales

Les mesures neurophysiologiques et comportementales en UX appliquée peuvent potentiellement répondre aux lacunes et insuffisances associées aux méthodes d'évaluation traditionnelles issues de l'utilisabilité, et représentent une nouvelle voie prometteuse dans l'élaboration de nouvelles méthodologies en expérience utilisateur. Les états émotionnels et cognitifs des utilisateurs peuvent être inférés en utilisant différents types de signaux physiologiques et comportementaux, tels que l'activité électrodermale, la fréquence cardiaque, l'oculométrie, l'analyse de la voix, les mouvements corporels, ou les expressions faciales (voir [16] et [17]).

Que ce soit dans le milieu universitaire ou dans des contextes industriels avancés, les signaux physiologiques utilisés lors d'une expérience doivent être choisis en fonction du construit psychologique d'intérêt. Ainsi, les signaux associés à la charge cognitive ont été utilisés dans ce mémoire. De plus, grâce aux technologies actuelles, toutes les techniques énumérées plus haut peuvent être plus facilement intégrées dans le coffre à outils des praticiens et chercheurs en UX.

L'électrocardiographie (ECG) est un outil de mesure de l'activité électrique des muscles du cœur, qui peut-être corrélée avec certains processus affectifs et cognitifs [18]. Par exemple, la fréquence cardiaque (HR) peut être corrélée à l'excitation, alors que la variation de la fréquence cardiaque (HRV) peut être associée à des changements de niveau de charge mentale ou la valence émotionnelle (positive ou négative) [19].

L'Activité électrodermale (EDA) désigne la variation des propriétés électriques de la peau en réponse à la sécrétion de la sueur par les glandes sudoripares eccrines [20]. L'activité électrodermale peut être corrélée à l'excitation et servir à mesurer les émotions [21] ainsi que la charge cognitive [22] au cours d'interactions avec un produit, système ou service.

La pupillométrie mesure la variation du diamètre de la pupille, qui répond non seulement aux changements de lumière ambiante (réflexe lumineux pupillaire), mais aussi à des stimuli cognitifs et émotionnels. Celle-ci répond aussi stimuli non visuels tels des pensées et des émotions [23]. De nombreuses recherches [24, 25] ont démontré que la dilatation de la pupille peut être liée à l'attention et à la charge cognitive, et ainsi fournir une estimation de l'intensité de l'activité mentale, et ce, que le participant soit conscient de ces changements ou non, et ce pendant toute la durée de l'expérience.

L'oculométrie est une technique consistant à enregistrer les mouvements des yeux et la position du regard de l'utilisateur, fournissant à la fois des mesures temporelles et spatiales [26]. Les fixations, arrêts momentanés de l'œil, et les saccades, les déplacements de l'œil d'une fixation à l'autre, donnent aux chercheurs et praticiens des indications sur l'information traitée par l'utilisateur durant une interaction humain-ordinateur [27].

Les émotions peuvent aussi être évaluées sur la base des expressions faciales, un des moyens les plus intuitifs de communication des états affectifs [28], grâce à des logiciels tels que FaceReader de Noldus [29] ou Affectiva [30]. Ces derniers sont des programmes d'analyse du visage, basés sur une nomenclature des mouvements des muscles faciaux issue des travaux d'Ekman [31], afin d'inférer la probabilité de sept émotions discrètes (la joie, la tristesse, la colère, la surprise, l'effroi, le dégoût et l'émotion neutre), en plus de la valence émotionnelle (négative par rapport au positif).

Il existe plusieurs autres types de mesures neurophysiologiques, telles que l'électroencéphalographie (EEG), enregistrement de l'activité électrique du cerveau; la respiration (RSP), essentielle à tous les systèmes physiologiques; et l'électromyogramme (EMG), qui mesure l'activité électrique générée par les muscles du visage. Cependant, ces mesures n'ont pas été utilisées dans le contexte de ce mémoire.

2.2.1. Avantages

Les mesures neurophysiologiques possèdent de nombreux avantages, tels que la temporalité et l'objectivité. En effet, ces mesures appliquées à l'UX permettront aux praticiens de recueillir des données quant à l'expérience vécue de l'utilisateur sans pour autant l'interrompre dans son interaction. La nature dynamique de ces signaux offre une fenêtre continue aux réactions des utilisateurs et peut fournir des informations précieuses sur ce qu'ils éprouvent durant l'interaction [15]. Ces signaux peuvent également être utilisés pour déceler des états émotionnels et cognitifs, dont l'utilisateur lui-même ignore l'existence ou dont il ne se souvient pas [32]. En outre, comparativement aux méthodes traditionnelles, ces méthodes ne présentent aucun biais rétrospectif [33].

De plus, l'utilisation de mesures physiologiques, en combinaison avec les méthodes d'évaluation traditionnelles, pourrait aider les praticiens à mieux mesurer l'expérience utilisateur, puisqu'ils fournissent chacun des informations complémentaires sur les émotions ressenties, les perceptions et les réactions des utilisateurs dans le contexte d'une interaction humain-ordinateur, soit avec un système, un jeu ou une interface Web. Alors que les méthodes d'évaluation traditionnelles peuvent offrir des données épisodiques, c'est-à-dire avant ou après l'interaction, les mesures physiologiques peuvent fournir de l'information continue, en temps réel [10]. L'ajout de mesures physiologiques pourrait aider les chercheurs et praticiens à identifier les réactions cognitives et émotionnelles des utilisateurs, tandis qu'un entretien après-tâche pourra les aider à approfondir leurs résultats.

2.2.2. Lacunes

Malgré les nombreux avantages associés aux mesures neurophysiologiques, l'utilisation de ces méthodes présente aussi certains inconvénients. En effet, la complexité de l'analyse des données nécessite souvent l'intervention d'experts, ce qui représente d'importantes contraintes de temps et d'argent; les coûts associés au traitement, à l'analyse et à l'interprétation de certains signaux représentant d'importants investissements de capital. De plus, ces méthodes nécessitent souvent des connaissances avancées dans diverses disciplines[3].Certains aspects techniques liés à l'utilisation des mesures physiologiques, par exemple, la pose des capteurs, la collecte et l'extraction de données doivent être exécutés par des experts possédant des connaissances en psychologie, en informatique ainsi qu'en statistique. Pour les praticiens et chercheurs novices, l'ajout de ces méthodes à leur coffre à outils représente une courbe d'apprentissage élevée.

[3] http://www.allaboutux.org/

Nonobstant, l'obstacle principal à l'utilisation des signaux physiologiques et comportementaux en UX demeure la contextualisation de ceux-ci ; en d'autres termes, leur valeur informative réduite lorsqu'ils ne peuvent être spécifiquement associés à un comportement ou état d'interaction donné [10, 18]. En effet, bien que nous puissions identifier les émotions qu'un utilisateur a ressenties durant son interaction, et ce à la seconde près, il est encore difficile d'identifier quel élément de l'interface a provoqué cette réaction, négative ou positive.

2.3. Objectifs de recherche

L'objectif de ce mémoire est de répondre à la question suivante : comment faire pour atténuer les barrières à l'entrée associée à l'utilisation des mesures physiologiques et ainsi en faciliter leur adoption dans le domaine de l'UX. Pour tenter de résoudre cette problématique, nous avons développé un outil d'évaluation de l'expérience utilisateur qui vise à contextualiser les signaux neurophysiologiques des utilisateurs, en plus de faciliter leur interprétation par l'entremise de cartes de chaleur physiologiques. Les trois articles de ce mémoire font état de deux collectes de données qui ont été réalisées afin de développer et évaluer notre outil. La première étude, réalisée en mai 2015, avait deux objectifs : 1) évaluer la capacité des cartes de chaleur physiologiques à mettre en évidence différents niveaux d'émotions et de charge cognitive expérimentés sur une même interface et 2) présenter un cas d'utilisation afin d'illustrer l'utilité de l'outil proposé pour l'évaluation de l'expérience utilisateur. La seconde collecte de données, effectuée en septembre 2016, avait pour but de valider le potentiel d'adoption de notre outil avec des experts de l'industrie de l'UX.

Présentation de l’article 1

Cet article présente notre réponse à la question suivante : comment permettre aux chercheurs et praticiens de l’UX de mesurer adéquatement les comportements et les états cognitifs et émotionnels des utilisateurs? L’intégration des mesures physiologiques représente, selon nous, une solution à ce problème. Cet article propose donc un nouvel outil d’évaluation de l’expérience utilisateur qui vise à contextualiser les signaux physiologiques et comportementaux des utilisateurs, afin d’atténuer les barrières à l’entrée associées à l’utilisation des mesures physiologiques. La charge cognitive est utilisée comme exemple de cas, afin de décrire l'outil proposé

Ce court article a été soumis et accepté en tant que travail en cours dans le cadre de la conférence Gmunden Retreat en NeuroIS 2015.

Chapitre 1
L'Idéation

Measuring visual complexity using neurophysiological data

Vanessa Georges, François Courtemanche, Sylvain Sénécal, Thierry Baccino, Pierre-Majorique Léger and Marc Fredette

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.

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].

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.

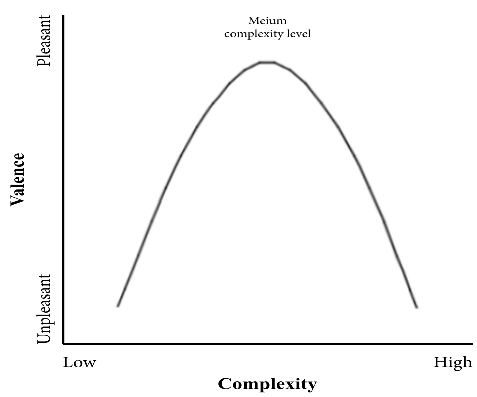


Figure 1 — The relationship between visual complexity and affective valence.

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.

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).

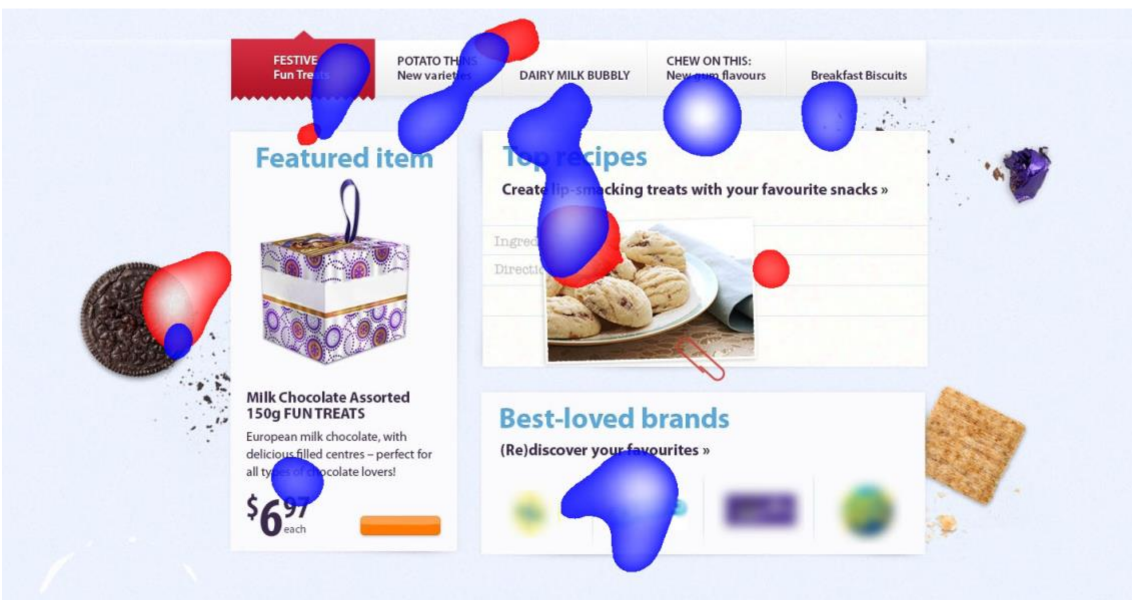


Figure 2 — Standard gaze heatmap versus a physiological heatmap based on pupil size. The blue gradient represents a standard gaze heatmap and the red gradient represents an arousal physiological heatmap based on pupil size.

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).

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.

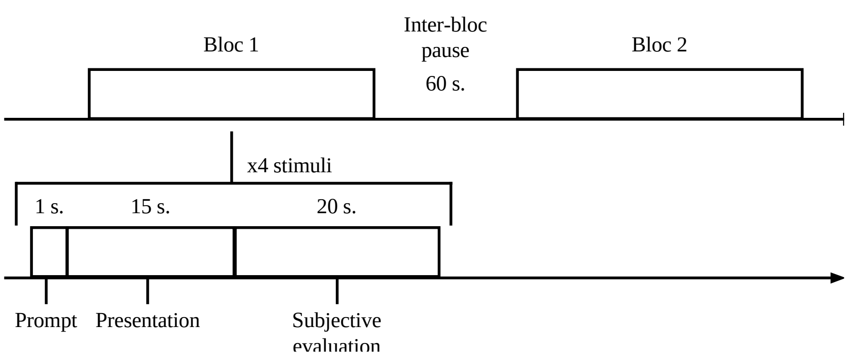


Figure 3 — Experimental design. Four images will be presented within each condition block. Blocks will be separated by a 60 seconds rest period to ensure the return of physiological signals to baseline level.

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.

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., N. Gauducheau, E. Jamet, G. Rouxel, and P. Salembier, La prise en compte des affects dans le domaine des interactions humain-machine. Le travail humain, 2011. 74 : p. 183-201.

3. Mahlke, S. and M. Minge, Consideration of Multiple Components of Emotions in Human-Technology Interaction, in Affect and Emotion in Human-Computer Interaction, C. Peter and R. Beale, Editors. 2008, Springer Berlin Heidelberg. p. 51-62.

4. Hassenzahl, M., The Interplay of Beauty, Goodness, and Usability in Interactive Products. Human–Computer Interaction, 2004. 19(4) : p. 319-349.

5. Tuch, A. N., J.A. Bargas-Avila, K. Opwis, and F. H. Wilhelm, Visual complexity of websites : Effects on users' experience, physiology, performance, and memory. International Journal of Human-Computer Studies, 2009. 67(9) : p. 703-715.

6. Michailidou, E., S. Harper, and S. Bechhofer. Visual complexity and aesthetic perception of web pages. in Proceedings of the 26th annual ACM international conference on Design of communication. 2008. ACM.

7. Ngo, D.C.L., L.S. Teo, and J.G. Byrne, Modelling interface aesthetics. Information Sciences, 2003. 152 : p. 25-46.

8. Oliva, A., M.L. Mack, M. Shrestha, and A. Peeper. Identifying the perceptual dimensions of visual complexity of scenes. in 26th Annual Cognitive Science Society. 2004. Austin, TX.

9. Berlyne, D., Studies in the New Experimental Aesthetics. 1974, Washington, DC : Hemisphere Publishing.

10. Courtemanche, F., P.-M. Léger, M. Frédette, S. Sénécal, and A. Dufresne, Method and Product for Visualizing the Emotions of a User, in Provisional patent application. 2015. p. 14.

11. Jennings, J.R., T. Kamarck, C. Stewart, M. Eddy, and P. Johnson, Alternate Cardiovascular Baseline Assessment Techniques : Vanilla or Resting Baseline. Psychophysiology, 1992. 29(6) : p. 742-750.

12. Wang, Q., S. Yang, M. Liu, Z. Cao, and Q. Ma, An eyetracking study of website complexity from cognitive load perspective. Decision Support Systems, 2014. 62 : p. 1-10.

13. Harper, S., E. Michailidou, and R. Stevens, Toward a definition of visual complexity as an implicit measure of cognitive load. ACM Transactions on Applied Perception (TAP), 2009. 6(2) : p. 10.

14. Laeng, B., S. Sirois, and G. Gredebäck, Pupillometry. Perspectives on Psychological Science, 2012. 7(1) : p. 18-27.

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, 1996. 93(5) : p. 1043-1065.

16. Parasuraman, R., Neuroergonomics brain, cognition, and performance at work. Current directions in psychological science, 2011. 20(3) : p. 181-186.

Présentation de l'article 2

Cet article présente la conception et le développement de notre outil d'évaluation de l'expérience utilisateur qui vise à contextualiser les émotions et construits psychologiques ressentis par les utilisateurs durant leur interaction avec un système. De plus, l'article fait état de la collecte de données ayant servi à son développement. Afin de faciliter la contextualisation et l'interprétation des données physiologiques, des heatmaps sont utilisés pour cartographier les différents états émotionnels et cognitifs des utilisateurs lors de l'interaction sur l'interface évaluée. Les heatmaps physiologiques sont une nouvelle méthode de visualisation qui représente l'intensité relative d'un état affectif ou cognitif sur une interface en utilisant des gradients de couleurs.

Cet article a été soumis et accepté dans le cadre de la conférence CHI — Computer-Human Interaction 2016

Chapitre 2
La conception

UX Heatmaps: Mapping User Experience on Visual Interfaces

Vanessa Georges, François Courtemanche, Sylvain Sénécal, Thierry Baccino, Marc Fredette and Pierre-Marjorique Léger

Abstract

In this paper, we present an off-the-shelf UX evaluation tool which contextualizes users' physiological and behavioral signals while interacting with a system. The proposed tool triangulates users' gaze data with inferred users' cognitive and emotional states to produce user experience (UX) heatmaps, which show where users were looking when they experienced specific cognitive and emotional states. Results show that for a given cognitive state (i.e., cognitive load), the proposed UX heatmap was able to effectively highlight the areas where users experienced different levels of cognitive load on an interface. The proposed tool enables the visual analysis of users' various emotional and cognitive states for specific areas on a given interface, and also to compare users' states across multiple interfaces, which should be useful for both UX researchers and practitioners.

Keywords: User experience; interface design; heatmaps; eyetracking; physiological computing; cognitive load; affective computing.

Introduction

High quality user experience (UX) has become a key competitive factor for product development [17]. User experience methods investigate how people feel about a system, game, or web interface, as opposed to how easily they managed to accomplish the task at hand, shifting the focus towards user affect. Assessing the range and quality of affects experienced while interacting with a system is now seen as critical for the development of products that satisfy both users' needs and expectations [14]. Measuring the emotional state of users during the interaction is therefore essential to the design of richer user experience. Traditional evaluation methods (e.g. questionnaires and interviews) rely on self-reported data in order to assess the affective and cognitive states of users, which are often exposed to different response effects, such as social desirability [18]. More so, these methods assess user perceptions either after or during the interaction, which induces retrospective biases or disrupts the user experience [9, 27].

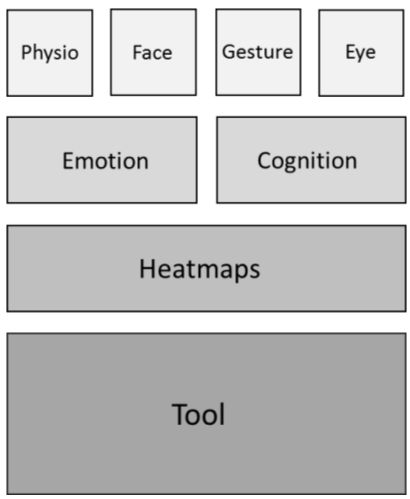


Figure 1 — From signals to UX heatmap tool.

Users' emotional and cognitive states can also be inferred using many different types of physiological and behavioral signals, such as electrodermal activity, heart rate, eyetracking, vocal and visual cues, body gestures, or facial expressions (see [37] and [6] for reviews). The dynamic nature of these signals offers a continuous window to the users' reactions, and can provide valuable insights as to what they are experiencing during the interaction. These signals can also be used to infer emotional and cognitive states, some of which the user himself is either unaware of or cannot recall when asked using traditional deferred methods [34]. Although physiological and behavioral measures are used in academia, and to a lesser extent in industrial contexts, many challenges remain [32]. These methods are often costly and require expert knowledge. However, the main obstacle to the use of physiological and behavioral signals remains their reduced informative value when they are not specifically associated with user behavior or interaction states [11, 25].

To meet this challenge, most researchers have concentrated their efforts on finding ways to measure physiological signals and interaction states synchronously. For example, Kivikangas et al. [19] have developed a triangulation system to interpret physiological data from video game events. Dufresne et al. [10] have proposed an integrated approach to eyetracking-based task recognition as well as physiological measures in the context of user experience research. Other researchers have also developed tools that allow users' to manually assign subjective emotional ratings on visual interfaces [16] or to visualize emotional reactions in terms of GUI widgets [7]. While these research streams have produced interesting results, they are not easily transferable to new contexts of use, as they are based on internal information from the interactive system (e.g., video game logs, application events, or areas of interest).

This paper presents a novel off-the-shelf and easy to interpret UX evaluation tool that aims to contextualize users' signals while interacting with a system interface. As illustrated in Figure 1, different emotional (sadness, happiness, surprise, etc.) and cognitive (cognitive load, stress, etc.) states are first inferred from continuous physiological or behavioral signals. The states are then triangulated with gaze data and mapped onto the user interface in order to create heatmaps and highlight areas where they occur with a higher frequency. The tool operationalizes this method by implementing different types of heatmap visualizations and evaluation functionalities. In order to describe the proposed tool, this paper focuses on the visualization of users' psychophysiological states using physiological signals. Cognitive load is used as the case example.

The remainder of the paper is organized as follow. First, a detailed account of the creation of physiological heatmaps is followed by an overview of the proposed tool.

Experimental validation results are presented. Applications of the tool to UX evaluation is discussed in the form of a use case using experimental data. Finally, the genericity of the tool is illustrated with the use of emotional heatmaps based on facial expressions.

Physiological Heatmaps

Traditional gaze heatmaps are used in eyetracking as intuitive representations of aggregated gaze data [22]. Their main use is to help researchers and HCI experts answer the question: "Where in the interface do people tend to look?" [35]. In the proposed visualization method, the users' gaze now serves as a mean of mapping physiological signals onto the user interface. The resulting heatmaps represent the physiological signals' distribution over the interface, and can help answer the following question: "Where in the interface do people tend to emotionally or cognitively react more strongly?" Below, we describe the four steps involved in the creation of physiological heatmaps: inference, normalization, accumulation, and colorization.

Inference

The regulation of emotional and cognitive states relies at once upon the sympathetic and parasympathetic activity of the autonomic nervous system, and thus requires physiological adjustments stemming from multiple response patterns [20]. As such, the relation between physiology and psychological states is more realistically described as a many-to-many relationship (i.e., multiple psychological states linked to multiple physiological variables) [5]. For example, when interacting with an interface, a single user's physiological signal (e.g., heart rate) can be associated with a change in cognitive load or emotional arousal. Therefore, the objective of the first step of the creation process is to disentangle this effect and ensure that the rendered heatmap truly represent the psychological construct of interest (i.e., cognitive load in the context of this paper). To do so, a machine learning classifier is used to estimate users' cognitive load.

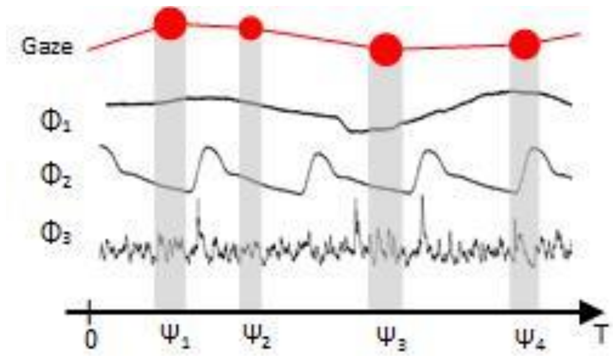


Figure 2 — Gaze-based psychophysiological inference process. Red dots represent eye fixations and red lines represent saccades. Φ1 = electrodermal activity, Φ2, = blood volume pressure, and Φ3 = pupil size. Ψx represent different inferences of a given construct of interest (e.g., different inferences of cognitive load).

As illustrated in Figure 2, all signals (Φ1, Φ2, Φ3) are segmented according to users' gazes and serve as features in the classifier. As each physiological system operates in collaboration with a variety of inputs and outputs from the rest of the organism, the measured signals present various durations and latencies for a given stimulus. For example, heart rate may have a shorter latency than electrodermal activity for a given stimulus. Therefore, each signal is segmented using a specific extraction window of different duration and latency starting at fixation onset. Latency is defined as the time elapsed between a fixation onset and the beginning of the extraction window and duration is defined as the time elapsed between the start and the end of the window. The identification of optimal latencies and durations, as well as the training of the classifiers are done using an empirical optimization process. For this study, the windows' optimization was done using a subpart of the data collected in the experiment. A detailed description of the segmentation and machine learning processes is given in [8]. This paper focuses on the heatmap visualization aspect of the tool. For each user's fixation, the inference outcome (Ψx) is used in the next step of the creation process.

Normalization

As physiological signals are subject to significant interpersonal variations or instrumental inaccuracies, absolute values cannot be used to compare data from multiple users. Physiological signals need to be corrected to account for the user's baseline [31]. In the proposed approach, the results of the physiological inference (see inference section) are normalized using z-score with the following equation:

$$W'_i = \frac{W_i - \mu}{\sigma}$$

where μ and σ are respectively the mean and the standard deviation of the inferred values for all of a user's fixations. This step also helps distinguish the "physiologically significant" areas of an interface from neutral ones. In traditional gaze heatmaps, every fixation has a positive intensity value and increases the height map. In our approach, physiologically unimportant fixations ($W' < 0$) are not considered in the subsequent accumulation step; only important physiological activity makes a contribution.

Accumulation

For traditional gaze heatmaps, the accumulation step consists in the creation of a blank map with the same dimensions as the image stimulus (n x m pixels). For each eye fixation, all the pixels are attributed an intensity level corresponding to an eyetracking metric (fixation count, absolute fixation duration, relative fixation duration, or participant percentage) multiplied by a scaling function (e.g., Linear, Gaussian), taking into account the distance between the gaze and the pixel [15]. Intensity values from various users falling on the same pixel are then summed to produce a height map representing the gaze intensity distribution over the image stimulus (Figure 3).

In the proposed method, the inference outcome (Ψx) is used to create the aforementioned height map, instead of an eyetracking metric. Therefore, the resulting height map represents the relative intensity of the physiologically inferred psychological state over the interface. Furthermore, as for standard gaze heatmaps, the height map is not rendered in its entirety. In order to outline the most prominent parts ofan interface, a parameter t ε [0, 1] is defined to compute the threshold under which intensity values are neglected. For example, t = 0.2 implies that only pixels with an accumulated intensity superior to 20% of the maximum intensity will be rendered in the colorization step. The value of the t threshold can be seen as representing the water level in a flooded valley. In this study, heatmaps were generated using Gaussian scaling and an intensity threshold of 0.8.

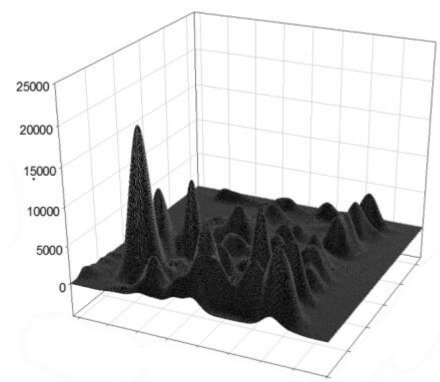


Figure 3 — Height map representation of aggregated gaze data.

Colorization

The last step in creating a heatmap is colorization. The main idea is to overlay on the stimulus image a semi-transparent layer that reflects the height map's variations. Height variations can be mapped to different color properties using a colorization function, resulting in various types of visualizations. The most commonly used visualizations are rainbow heatmaps, luminance maps, and contrast maps [15]. Breslow et al. [4] have demonstrated that multicolored scales (e.g., rainbow) are best suited for identification tasks (i.e., determining absolute values using a legend) and single hue scales (e.g., luminance) are best suited to compare relative values.

Therefore, in this work, the colorization step uses a luminance gradient (i.e., an increase in color brightness is associated with an increase in the height map). Furthermore, the use of single hue gradients allows the mapping of multiple constructs at the same time.

Cross-stimuli Colorization

The proposed tool also features a novel colorization function : cross-stimuli gradient. This colorization function displays a single gradient across multiple stimuli, in order to evaluate users' experience of one interface relative to others.

Eyetracking analysis software provide heatmap representations on a single image basis only. Heatmaps are generated one at a time and the full gradient (i.e., all possible color properties increments) is displayed on each image. The colorization function associates pixels under the maximum height with the maximum value of the color gradient, and pixels at the threshold height with the minimum value. For example, when generating a rainbow heatmap on a given interface, a gradient going from red to green, by way of yellow and orange, is mapped in its entirety onto the image. In a luminance gradient, the maximum value of the color gradient is the absence of hue, i.e. white. This implementation is relevant for gaze heatmaps as they are meant to represent how a specific amount of aggregated eye fixations is distributed within a single interface. However, it may be interesting to evaluate the user's experience of said interface relatively to different interfaces.

To do so, we propose a colorization function that uses a cross-stimuli gradient. First, the height maps of a specific set of image stimuli that are to be analyzed together are merged together. Second, the colorization function associate pixels' height of each rendered heatmap to a gradient level, relatively to its location in the common height map. Therefore, the full gradient representing the user experience construct spans across all images and, allows users to compare areas of different interfaces together. Unlike standard gaze heatmaps, heatmaps rendered using this cross- stimuli gradient do not necessarily present hot spots (i.e., doesn't use the full gradient within the same interface). Hot spots are present only on interfaces that present the highest density of the visualized user's state compared to the others. For example, figure 5 illustrates three interfaces representing various levels of visual complexity (low, medium and high). Instead of comparing three individually constructed heatmaps (i.e. one per image), a single cross-stimuli gradient is mapped across all three stimuli. The hot spot is located on the top left corner of the third stimulus, meaning that users experienced the highest level of cognitive load in this specific area, relative to the other versions of the interface.

UX Heatmap Tool Description

The evaluation tool enables the contextualization of various physiologically inferred affective and cognitive states, during users' interaction with an interface Figure 4 illustrates the data processing sequence required to generate heatmaps.

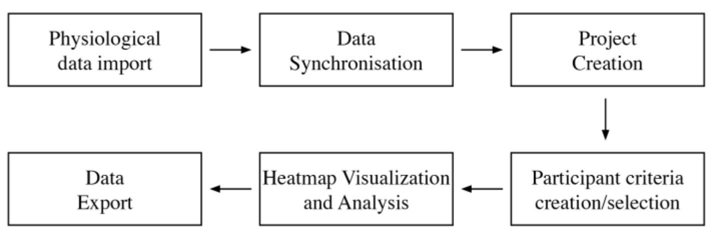


Figure 4 — Data process sequence.

First, data files from various physiological and behavioral recording devices are imported and synchronized. Upon project creation, participant selection (e.g., women only) is required, followed by heatmap creation and rendering. Multiple heatmaps can be displayed simultaneously (e.g., negative valence and cognitive load, or negative and positive valence). The opacity and display color of each heatmap can be individually selected, before or after visualization, in order to enable the comparison of multiple emotional and cognitive constructs. The resulting heatmaps can then be exported in jpeg format.

Participant Selection

Participant data can be created and managed before or after the creation of a new project. This data consists of an identifier and variables, such as gender and age. These can be used to create participant groups; which users can subsequently use to filter the physiological data when producing heatmaps. These variables can be selected using a drop-down menu on the left hand panel. All participants are displayed in the participant list (see Figure 5).

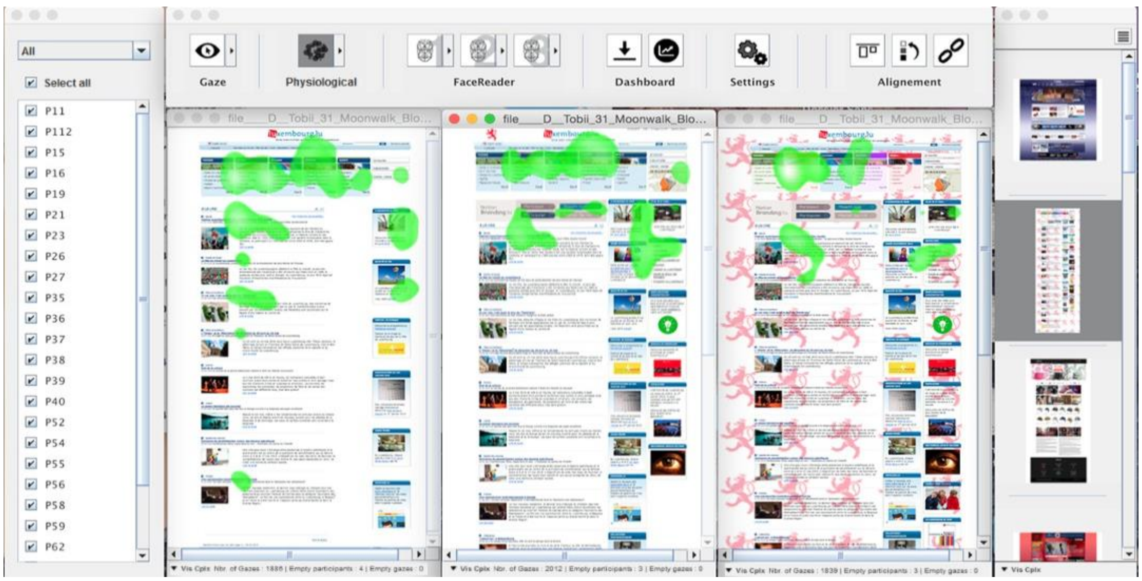


Figure 5 — Tool overview. Participant data can be created and managed using the left hand panel. Images used during a test or experiment are displayed on the right side panel. Above: An example of a cross-stimuli gradient is illustrated using images representing three levels of cognitive load (from left to right: low, medium and high).

Construct Manager

Three types of heatmaps can be generated in the current state of the tool: gaze heatmaps, physiological heatmaps (based on cognitive load), and facial expressions heatmaps (emotion). Emotion heatmaps are generated using the FaceReader 6 software (Noldus, Netherland), which infers the probability of seven discrete emotions (happy, sad, angry, surprised, scared, disgusted and neutral) and emotional valence (negative vs. positive), based on facial movements [36]. Given that FaceReader recordings are synchronized with eyetracking data, emotion prediction is introduced in the accumulation step of heatmap creation (see physiological heatmaps section). As seen in Figure 5, up to three emotions can be mapped onto the interface at any given time during the analysis, to allow the simultaneous comparison of various emotional states. Heatmaps can all be displayed one by one or simultaneously based on user selection. Participant selection and group filters are applied to all visualizations.

Interface Selection

Images used during the analysis are displayed on the right side panel. Users can select as many stimuli as desired. Upon heatmap visualization, information relative to selected stimuli (e.g., number of gazes) will appear at the bottom of the prompted window.

Experimental Validation

A lab experiment was conducted to develop and evaluate the proposed tool. In this validation study, cognitive load was used as the psychological construct of interest. The goal of the evaluation was twofold: 1) assess the ability of the physiological heatmaps method to effectively highlight the different levels of cognitive load experienced on an interface, and 2) present a use case in order to illustrate the usefulness of the proposed tool for user experience evaluation. Two separate tasks were designed to achieve these goals.

Participants

For this experiment, a total of 44 students between the ages of 18 and 35 were recruited through the student panel at HEC Montréal over a period of five weeks. Data from 18 participants were rejected due to equipment malfunction, data synchronization imprecision, or insufficient eyetracking calibration precision. Therefore, data from 26 participants were used in the analyses, from which 17 were female, for an average age of 24. Participants had normal or corrected- to-normal vision and were pre-screened for glasses, laser eye surgery, astigmatism, epilepsy, and neurological and psychiatric diagnoses. The total experiment duration was of one hour, during which participants were asked to perform the two experimental tasks. Compensation in the form of a 20$ gift certificate was given to each participant upon completion of the experiment.

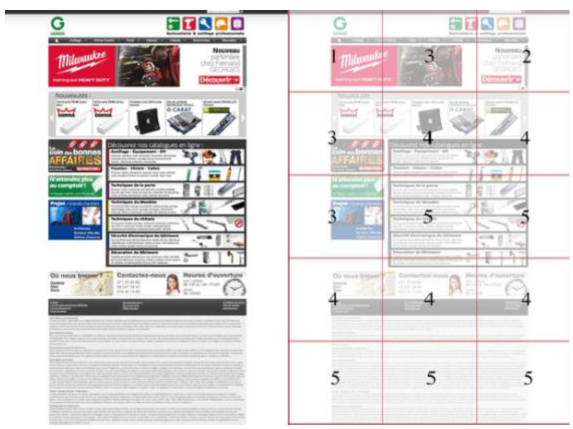


Figure 6 — Evaluation task stimuli. Left: Stimulus 7 was rated as medium complexity website, with inter-judged average visual complexity rating of 4.85 out of 10. The task pertaining to this particular stimulus, given to evaluating judges and subjects alike was to search for information regarding parts and equipment brands available in store. Right: Evaluation sheet for Stimulus 7 with an example of a participant's ratings.

Stimuli

In the experiment, the visual complexity of website homepages was manipulated in order to trigger different cognitive load levels. Visual complexity is closely related to the perceived cognitive load users think will be required to interact with an interface [13] and has also been shown to be correlated with cognitive load [33]. Multiple studies have aimed at understanding how the cognitive system interprets different levels of image complexity (see [37], [6] and [26]). Most studies consider that there are three types of cognitive load (CL): intrinsic load, extraneous load, and germane load. According to Sweller [30], intrinsic cognitive load is linked to the material's content, extraneous cognitive load is based on the presentation forms, and germane cognitive load involves information consolidation. Website complexity relates to extraneous cognitive load [31]. Wang et al. [31] also found that extraneous CL can be decreased by adequate visual presentation and design of material.

These images were presented in an Internet browser, as to maintain the aspect and the original length of the homepage. Interaction with homepages was limited to scrolling. Websites were selected to reflect a wide range of tasks and sectors, such as business, academia, entertainment and leisure. In order to eliminate content-related affect, news websites were excluded. Local and national websites were also excluded in order to eliminate any bias based on familiarity.

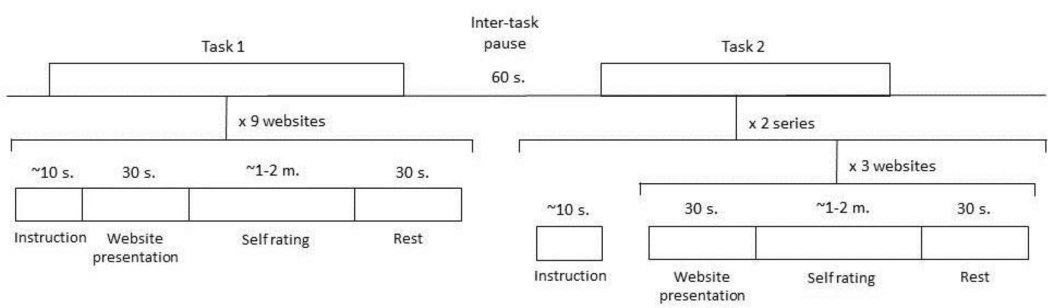


Figure 7 — Experimental procedure.

Task 1

The goal of the first experimental task was to gather a physiological data set representative of different levels of experienced cognitive load, in order to assess the accuracy of the proposed method. The stimuli included nine homepages representing three levels of visual complexity: 3 low, 3 medium, and 3 high.

In order to select these nine homepages, twelve undergraduate students were instructed to rate the overall visual complexity of twenty-one website homepages, as defined as the degree of difficulty in understanding the information presented. The complexity scores' agreement was calculated using a two-by-two correlation procedure. Agreement between all 66 possible pairs of judges was averaged for each individual homepage. The resulting average correlation (for all homepages) was of $r=0.601$ ($p=0.047$). The stimuli were presented as illustrated in Figure 7 (Task 1).

At the beginning of each trial, participants were given a task to perform on the forthcoming homepage. They were instructed to look at the homepage having in mind how they would execute the task (e.g., search information about products and brands available in store). The website was then presented in a web browser for 30 seconds. Following the presentation, a printed version of the same interface was presented for an unlimited amount of time, featuring a red 15 box grid (see Figure 6). Participants were asked to assess the visual complexity of each of the 15 regions of the homepage using a 1 to 5 rating scale, 5 being very complex.

Task 2

The objective of the second task was to illustrate the usefulness of the proposed tool for webpage prototype comparison. Two series of three website homepages were doctored in order to reflect three different level of perceived visual complexity: low, medium, and high (Figure 5). Starting from a single homepage, two duplicate versions of the images were created. The interface was thus simplified (low complexity), as well as rendered more complex (high complexity). According to Oliva et al. [23], visual complexity is mainly represented by the perceived dimensions of quantity of objects, clutter, openness, symmetry, organization, and variety of colors. These elements were used to alter the complexity level of the images. The content was not altered. The images were evaluated by undergraduate students, in order to assess their visual complexity. They were modified and re- evaluated as need be. As illustrated in Figure 7 (Task 2), participants were first instructed on the task they would have to execute. The three versions of the website were then presented with a self- rating and rest period. The presentation order was randomized for each participant. The same procedure was repeated with another series of websites. Task 2 lasted twenty minutes on average.

Physiological Signals and Equipment

The signals used for physiological heatmaps should be selected according to the psychological construct of interest. Thus, signals known to be related to cognitive load were selected in this study.

A Biopac MP150 amplifier (Biopac MP) was used to record two peripheral physiological signals: electrocardiogram (ECG) and electrodermal activity (EDA). ECG-derived measures have been shown to correlate with affective and cognitive processes. For example, heart rate can be associated with shifts from low to high mental workload [12]. EDA measures the activity of the eccrine sweat glands and has been shown to be correlated to arousal. It can be used to measure emotions [3] and cognitive load [29] during system interactions. EDA was recorded using two electrodes placed on the palm of the non-dominant hand. Both physiological signals were recorded with a sampling rate of 500Hz.

Eye fixations, pupil diameter, and blink rates were recorded with a Tobii X-60 eyetracker (Tobii Technology AB). Research shows that variations in pupil size respond significantly to cognitive and emotional stimuli [21]. Ambient illumination was controlled and kept constant during the experiment in order to minimize pupil size variations due to light reflex. Blink rate can be affected by external stimulus or by emotional and cognitive states such as fatigue, but has been shown to be a perceptual load indicator (where information can be noticed or unnoticed) in the context of cognitive load measurement [1]. A nine point calibration was performed for all participants, and was repeated until sufficient accuracy was achieved. Current video-based eye trackers have a spatial resolution of up to 0.01o/ 2 kHz [15]. However, such a high resolution level is experimentally hard to achieve due to subjects' variability. In this research, sufficient accuracy was defined as ~ 1 cm around the center of the calibration points, which led to participants' dismissal, if not obtained. Stimuli were presented on a 22' LG LED monitor with a resolution of 1680 x 1050 pixels and a refreshing rate of 60Hz.

Videos of the participants' face were recorded using a webcam and the Media Recorder 2 software (Noldus, Netherland). Videos were processed in FaceReader 6 (Noldus, Netherland) to produce emotional states inference.

Results

The effectiveness of the proposed method to predict spatial locations of greater cognitive load has been tested using data from Task 1. Data from Task 2 is analyzed qualitatively in the Discussion section. Data was analyzed in order to evaluate the capacity of physiological heatmaps to capture experienced cognitive load variance over the different webpages. As a relative baseline, correlation with gaze heatmaps was also calculated. Although they are not meant to assess cognitive load, traditional gaze heatmaps represent the closest data visualization method to which we can compare physiological heatmaps. As illustrated in Figure 8, data from Task 1 was analyzed by comparing the highest peaks of the intermediate height maps with the underlying user ratings. One height map was generated per participant for each homepage.

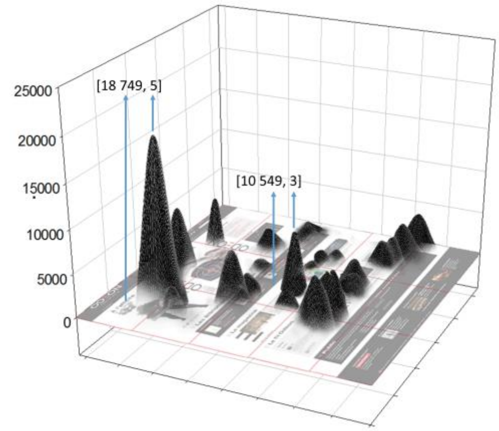


Figure 8 — Evaluation data example. The first self-rating (top- left) square has a value of 5 (on a maximum of 5) and the highest peak of the overlapping portion of the height map is of 18 749 (relative scale). One data point is generated per rating square for a maximum of 15 per website per subject. The evaluation square over which the height map is null are not used in the analyses.

This process expected a total of 3 510 data points (9 websites x 15 ratings x 26 subjects). However, 2 749 data points were obtained for physiological heatmaps and 2 147 data points for gaze heatmaps as only ratings over which the height map is not null were considered. Therefore, websites' areas that were not looked at during the 30s presentation (but rated afterwards) were not used in the analyses. As shown in Table 1, physiological heatmaps were significantly related to users' ratings of visual complexity (R^2=.291, p-value<.000). Using the distribution of the R^2 statistics [24], we can also affirm that the R^2 obtained using the physiological heatmaps is significantly higher than the one obtained with gaze heatmaps (z=3.024; p-value=0.02).

Visualization	R^2 (p-value)
Physiological heatmaps	0.291 (<0.000)
Gaze	0.167 (<0.000)

Table 1 — R^2 between highest peaks and users' ratings.

The p-values were corrected by assessing the effect of each measure on the ratings values by using linear regression model accounting for the potential correlation between each repeated measure coming from the same subject. These results show that cognitive load was successfully assessed by the proposed physiological heatmap, which can be complementary to traditional gaze heatmap. While gaze heatmaps can identify the “where”, UX heamaps can be used to address the “why” of users’ gaze behavior .

Discussion

User experience evaluation needs differ between industry and academia. While the former’s needs are to analyze and adequately communicate findings in order to improve UX, the latter’s interest resides in the validation and understanding of phenomena, based on hypothesis [28]. This section illustrates two types of analysis that can be undertaken with the described tool: 1) a comparative analysis of various prototypes using a cross-stimuli gradient and, 2) an exploratory analysis using emotional and cognitive state heatmaps.

Cross-stimuli Analysis

First, two sets of three images, reflecting three different levels of perceived visual complexity were created. In the first of the two series, participants were asked to find information about upcoming events that would interest them for a visit to Luxembourg. These images, presented in figure 5, were then analyzed using the cross-stimuli gradient.

When analyzing participants’ cognitive load heatmaps, we can see which areas of each interface induced higher levels of cognitive load. More so, when compared to the other images, we can also ascertain which interface induced greater overall cognitive load. Of the three images, the area which inferred greater cognitive load is located at the top left corner of the most complex image (Figure 5 - right-hand side). When comparing this area with the same areas within the other two images (i.e. located in the same spatial location), we can see that the information content has not changed. Yet, the experienced cognitive load in the most complex version of the interface differs from the others. This is in line with the manipulations performed on the aforementioned image. As mentioned in the Experimental Validation section, changes in color, including the hue, and number of colors, clutter and the quantity of objects were manipulated in order to render the image more complex. The cognitive load heatmaps indicate that these changes influenced participants' perceived visual complexity, therefore experiencing higher cognitive load. Simply put, the number of colors and elements surrounding the area made it more difficult for users to assimilate the information, requiring more cognitive resources.

Looking at Figure 5, again using the cross-stimulus gradient, we can also see that changes in visual complexity also impacted users' behavior. Fixations on the third webpage (right-hand side of Figure 5) seemed to be aggregated in specific locations, in the top half of the interface, when compared to the less complex images where fixations were more widespread (left-hand side). Participants seemed to focus their attention and efforts towards specific areas, when faced with a more complex interface. One of the conclusion we can draw from this visualization is that users' seemed hesitant to explore websites deemed more complex. As stated by Harper et al. 2009 [13], "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." This statement is in line with our analysis, as UX heatmaps show user's reluctance to explore the page, anticipating the required cognitive load needed to do so.

Figure 9 — Negative, positive and cognitive load heatmaps examples. On the left, negative valence (red) and positive valence (yellow) heatmaps are illustrated. On the right, negative valence (red) and cognitive load (green) heatmaps.

In the context of this task, the comparison of various images using a cross-stimuli gradient allowed us to see the effect of design changes on user experience, both in terms of behavior and experienced cognitive load. For practitioners, this visualization can help make a more informed decision in a prototype comparison or A/B testing context.

Analysis of Emotional and Cognitive States

The Circumplex model of affect describes emotions using the two dimensions of valence and arousal [35]. Valence is used to contrast states of pleasure (e.g., happy) and displeasure (e.g., angry), and arousal to contrast states of low arousal (e.g., calm) and high arousal (e.g., surprise). Using the second set of images, we compared regions of high cognitive load versus regions of negative emotional valence; as well as areas of negative and positive valence.

Figure 9 shows the simplified version of a popular e- commerce website used during Task 2, in which participants were asked to look at the page and select an item they would like to purchase. When looking at the left-hand side of Figure 9, we can see the mapping of a positive emotional valence heatmap (yellow) and a negative emotional valence heatmap (red). Higher intensity emotion areas, or hotspots, are not located on the same pictorial information when comparing the two visualizations. The negative valence heatmap indicates that users experienced displeasure with higher frequency on text and navigation areas, whereas the positive valence heatmap hotspots are located on the human faces at the top of the interface, as well as on the video game (bottom right corner).

The right-hand panel of Figure 9 represents the mapping of a cognitive load heatmap (green) and a negative emotional valence heatmap (red). As we can see, regions of high cognitive load and areas of negative valence overlap. Therefore, in the context of this task, we can observe that high cognitive load, induced by the visual complexity of the page, led users to experience negative emotions. Furthermore, these clustered fixations seem to be located predominantly on text heavy areas.

In this example, comparing regions of high cognitive load to regions of negative emotional valence can help UX practitioners to not only identify the problematic areas of an interface, for instance areas that can adversely affect the user in the completion of a task or that could serve as deterrent to the use of a system, but also contextualize these emotional states by highlighting the graphical elements behind such states.

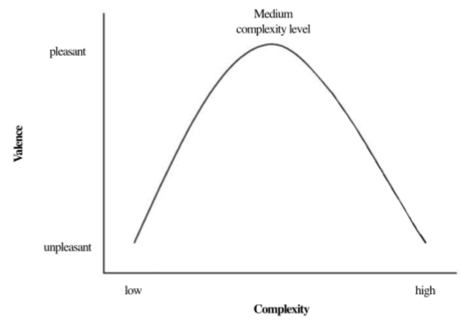


Figure 10 — The relationship between visual complexity and affective valence.

From a research perspective, such a tool can be used to visualize the effects of high cognitive load on users' emotional state. For instance, the tool's triangulation of UX constructs makes it possible to test the proposed inverted U- shaped relationship between complexity and emotional valence, proposed by Berlyne (1974) [2]. Simply put, interfaces either deemed too simple or too complex, will result in lower affective valence. As shown in Figure 10, results obtained in task 2 are in line with this finding as we can see that areas of high cognitive load overlap areas of negative valence.

Limitations and Future Works

The proposed tool aims to support the work of experts (i.e., ergonomists, designers, researchers) from the HCI and UX communities. While UX heatmaps can communicate data, they cannot make a diagnosis. To maximize their usefulness, the tool should be integrated to the methods already available to experts (e.g. questionnaires, observations, interviews). Furthermore, as for most eyetracking software, the proposed tool can only be used to analyze interactive interfaces at different points in time, i.e. different stills of the interface, but not interfaces in motion (e.g. flash animations). This represents the main limitation of the tool.

The next step for future work is the inclusion of other cognitive and emotional states, such as emotional UX heatmaps based on physiological signals, as opposed to FaceReader data in the tool's current version. Moreover, we will be conducting interviews with researchers and practitioners in order to evaluate the UX heatmap tool, and to better understand the needs of the UX community for future iterations.

Conclusion

The objective of this paper was to present an off-the-shelf, easy to interpret UX evaluation tool which contextualizes users' signals while interacting with a system. Using these signals to infer users' emotional and cognitive states and mapping these states using UX heatmaps on the interface provides researchers and practitioners with a useful tool to contextualize users' reactions. This triangulated approach makes it possible to visually analyze users' various emotional and cognitive states for specific areas of a given interface (e.g., cognitive load combined with emotional valence). Moreover, the cross-stimuli color gradient makes it possible to compare users' gaze patterns and states across multiple interfaces (e.g., A/B testing).

Acknowledgments

Authors want to thank Brendan Scully for manuscript revision. The authors also want to thank the research assistants who administered the study. This work was supported by the FRQSC (Fonds de Recherche du Québec — Société et Culture).

References

1. Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A. and Montanari, R. 2011. Driver workload and eye blink duration. Transportation Research Part F : Traffic Psychology and Behaviour 14, 3: 199-208.

2. Berlyne, D. Studies in the New Experimental Aesthetics. Hemisphere Publishing, Washington, DC, 1974.

3. Boucsein, W. Electrodermal Activity. Springer, Berlin, 2012.

4. Breslow, L.A., Ratwani, R.M. and Trafton, J.G. 2009. Cognitive Models of the Influence of Color Scale on Data Visualization Tasks. Human Factors: The Journal of the Human Factors and Ergonomics Society 51, 3: 321-338.

5. Cacioppo, J.T., Tassinary, L.T. and Berntson, G. Psychophysiological Science : Interdisciplinary Approaches to Classic Questions About the Mind. in Cacioppo, J.T., Tassinary, L.G. and Bernston, G.G. eds. Handbook of Psychophysiology, Cambride University Press, New York, 2007, 1-18.

6. Calvo, R.A. and D'Mello, S. 2010. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. Affective Computing, IEEE Transactions on 1, 1 : 18-37.

7. Cernea, D., Weber, C., Ebert, A. and Kerren, A., Emotion scents: a method of representing user emotions on gui widgets. in IS&T/SPIE Electronic Imaging, (2013), International Society for Optics and Photonics, 86540F-86540F-86514.

8. Courtemanche, F., Dufresne, A. and LeMoyne, É. Multiresolution Feature Extraction During Psychophysiological Inference: Addressing Signals Asynchronicity. in da Silva, H.P., Holzinger, A., Fairclough, S. and Majoe, D. eds. Physiological Computing Systems, Springer Berlin Heidelberg, 2014, 42-55.

9. de Guinea, A.O., Titah, R. and Léger, P.-M. 2014. Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation. Journal of Management Information Systems 30, 4 : 179-210.

10. Dufresne, A., Courtemanche, F., PromTep, S. and Sénécal, S., Physiological Measures, Eye Tracking and Task Analysis to Track User Reactions in User Generated Content. in Measuring Behavior, (Eindhoven, The Netherlands, 2010), 218-222.

11. Ganglbauer, E., Schrammel, J., Deutsch, S. and Tscheligi, M. Applying Psychophysiological Methods for Measuring User Experience: Possibilities, Challenges and Feasibility User Experience Evaluation Methods in Product Development (UXEM'09) workshop at the 12th IFIP TC13 Conference on Human-Computer Interaction, 2009.

12. Haapalainen, E., Kim, S., Forlizzi, J.F. and Dey,A.K. Psycho-physiological measures for assessing cognitive load Proceedings of the 12th ACM international conference on Ubiquitous computing, ACM, Copenhagen, Denmark, 2010, 301-310.

13. Harper, S., Michailidou, E. and Stevens, R. 2009. Toward a definition of visual complexity as an implicit measure of cognitive load. ACM Transactions on Applied Perception (TAP) 6, 2: 10.

14. Hassenzahl, M., Diefenbach, S. and Göritz, A. 2010. Needs, affect, and interactive products– Facets of user experience. Interacting with computers 22, 5 : 353-362.

15. Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R. and Jarodzka, H. Eye Tracking: A comprehensive guide to methods and measures. Oxford University Press, 2011.

16. Huisman, G., Hout, M.v., Dijk, E.v., Geest, T. v. d. and Heylen, D. LEMtool : measuring emotions in visual interfaces Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Paris, France, 2013, 351-360.

17. Kaasinen, E., Roto, V., Hakulinen, J., Heimonen, T., Jokinen, J.P., Karvonen, H., Keskinen, T., Koskinen, H., Lu, Y. and Saariluoma, P. 2015. Defining user experience goals to guide the design of industrial systems. Behaviour & Information Technology, ahead-of-print : 1-16.

18. King, M.F. and Bruner, G.C. 2000. Social desirability bias: A neglected aspect of validity testing. Psychology and Marketing 17, 2 : 79-103.

19. Kivikangas, M., Nacke, L. and Ravaja, N. 2011. Developing a triangulation system for digital game events, observational video, and psychophysiological data to study emotional responses to a virtual character. Entertainment Computing 2, 1 : 11-16.

20. Kreibig, S.D. 2010. Autonomic nervous system activity in emotion: A review. Biological Psychology 84, 3 : 394-421.

21. Laeng, B., Sirois, S. and Gredebäck, G. 2012. Pupillometry. Perspectives on Psychological Science 7, 1 : 18-27.

22. Nielsen, J. and Pernice, K. Eyetracking Web Usability. New Riders, Berkeley, California, 2012.

23. Oliva, A., Mack, M.L., Shrestha, M. and Peeper, A., Identifying the perceptual dimensions of visual complexity of scenes. in 26th Annual Cognitive Science Society, (Austin, TX, 2004).

24. Olkin, I. and Finn, J.D. 1995. Correlations redux. Psychological Bulletin 118, 1 : 155.

25. Pantic, M., Nijholt, A., Pentland, A. and Huanag, T.S. 2008. Human-Centred Intelligent Human-Computer Interaction (HCI2): How far are we from attaining it? International Journal of Autonomous and Adaptive Communications Systems 1, 2: 168-187.

26. Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin. 124, 3 : 372-422.

27. Robinson, M.D. and Clore, G.L. 2002. Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. Journal of Personality and Social Psychology 83, 1: 198-215.

28. Roto, V., Vermeeren, A.P.O.S., Väänänen-Vainio-Mattila, K., Law, E. and Obrist, M., Course notes : User Experience Evaluation Methods - Which Method to Choose? in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (Paris, France, 2013), ACM.

29. Shi, Y., Ruiz, N., Taib, R., Choi, E. and Chen, F. Galvanic skin response (GSR) as an index of cognitive load CHI « 07 Extended Abstracts on Human Factors in Computing Systems, ACM, San Jose, CA, USA, 2007, 2651-2656.

30. Sweller, J. 1988. Cognitive load during problem solving: Effects on learning. Cognitive science 12, 2 : 257-285.

31. van den Broek, E., van der Zwaag, M.D., Healey, J.A., Janssen, J.H. and Westerink, J.H.D.M. Prerequisites for Affective Signal Processing (ASP) — Part IV 1st International Workshop on Bio-inspired Human-Machine Interfaces and Healthcare Applications — B-Interface 2010, Valencia, Spain, 2010, 59-66.

32. Vermeeren, A.P.O.S., Law, E.L.-C., Roto, V., Obrist, M., Hoonhout, J., V, K., #228 and nen-Vainio-Mattila. User experience evaluation methods: current state and development needs Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, ACM, Reykjavik, Iceland, 2010, 521-530.

33. Wang, Q., Yang, S., Liu, M., Cao, Z. and Ma, Q. 2014. An eye-tracking study of website complexity from cognitive load perspective. Decision Support Systems 62 : 1-10.

34. Ward, R.D. and Marsden, P.H. 2003. Physiological responses to different WEB page designs. International Journal of Human-Computer Studies 59, 1-2 : 199-212.

35. Wooding, D.S. Fixation maps : quantifying eye-movement traces Proceedings of the 2002 symposium on Eye tracking research & applications, ACM, New Orleans, Louisiana, 2002, 31-36.

36. Zaman, B. and Shrimpton-Smith, T. The FaceReader: measuring instant fun of use Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, ACM, Oslo, Norway, 2006, 457-460.

37. Zhihong, Z., Pantic, M., Roisman, G.I. and Huang, T.S. 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 1 : 39-58.

Présentation de l'article 3

Afin de démontrer la pertinence de l'outil dans le domaine de l'expérience utilisateur, nous avons conduit une seconde collecte impliquant des chercheurs et praticiens du domaine. L'objectif de cette seconde collecte de données était d'obtenir des commentaires sur les manières de faciliter l'adoption des mesures physiologiques chez les praticiens de l'UX. Au cours de cette expérimentation, les sujets ont été amenés à commenter la pertinence et la facilité d'utilisation des fonctionnalités de l'outil. Cet article fait état de ces résultats.

Cet article a été soumis et accepté comme étude de cas dans le cadre de la conférence HCII — Human-Computer Interaction International 2017.

Chapitre 3
L'Évaluation

The Adoption of Physiological Measures as an Evaluation Tool in UX

Vanessa Georges, François Courtemanche, Sylvain Sénécal, Pierre-Marjorique Léger, Lennart Nacke and Romain Pourchon

Abstract

One of the challenges associated with the use of physiological signals as an evaluation tool in measuring user experience (UX) is their reduced usefulness when they are not specifically associated with user behavior. To address this challenge, we have developed a new evaluation tool which contextualizes users' physiological and behavioral signals while interacting with a system. We have conducted interviews with 11 UX practitioners, from various industries, to evaluate the usefulness of our tool. Through these interviews we gained a better understanding of the challenges facing industry practitioners when using physiological measures and assessed the functionalities provided by our tool.

Keywords: User experience; Interface design; Heatmaps; Eyetracking; Physiological computing; Cognitive load; Affective computing.

Introduction

User experience (UX) has recently become of strategic importance in the information technology industry [14]. The Tech3Lab is an applied research lab in human-computer interaction at the HEC Montréal business school, specializing in user experience using eyetracking and neurophysiological and behavioral measures. Our research pertains to the development of new evaluation methods, ones that investigate the why instead of the how as information on how users feel about a system, game, or web interface is now a common requirement for all UX evaluation methods [1].

Our recent work with our industry partners has lead us to question a major discrepancy between industry and academic practices: while physiological measures are increasingly used in academia, the adoption of these methods as UX evaluation tools remains uncommon in industry. We have observed a growing demand for more quantitative user research to provide data-driven recommendations for change, which we implement using eyetracking and neurophysiological and behavioral measures. We therefore wanted to understand what can be done to facilitate their adoption in industry. In tackling this issue, we have sought to create a visualization tool that contextualizes physiological and behavioral signals to facilitate their use [4]. The visualization method that we created is UX heatmaps, an integrated visualization tool which contextualizes physiological and behavioral signals to facilitate the interpretation of these measures [12].

Physiological Measures in UX

Traditional evaluation methods other than direct observation, for example questionnaires or interviews, mostly rely on self-reported data to assess the affective and cognitive states of users either during or after the interaction [6]. For example, Hassenzahl et al have developed a questionnaire to evaluate users » feelings about a system [11]. The results assess the user's reflection on the interaction, but not the interaction itself [13]. Users' emotional and cognitive states can also be inferred using physiological signals, such as electrodermal activity, heart rate, eyetracking and facial expressions (see [2] and [3] for reviews). As an evaluation method, electrodermal activity (EDA), which measures the electrical conductance of the skin, can provide practitioners with real-time information as to what the user is experiencing throughout the interaction. EDA is used as an indication of physiological arousal [8], as well as emotions. FaceReader [7], which analyzes facial expressions and infers the probability of seven discrete emotions (happy, sad, angry, surprised, scared, disgusted and neutral) and emotional valence (negative vs. positive) based on facial movements, can provide important temporal information without retrospective or social desirability bias. Furthermore, data is collected without interrupting the user in their authentic interaction.

However, these measures are still difficult to contextualize and interpret, as they are not specifically associated with user behavior or interaction states. Let's take the example of a user asked to browse the product offerings of an e-commerce website and purchase an item. With physiological data, we can infer that the user was frustrated at some point during the interaction, for example during the checkout process, but not the element that caused the negative emotion. We are therefore left wondering what was the button, task or area of the interface which caused the user to feel frustrated or angry. Physiological signals also require a certain degree of interpretation, as the output needs to be processed to transition from raw data to useful actionable insights. To meet these challenges, Kivikangas et al. [15] have developed a triangulation system to interpret physiological data from video game events. Other researchers have also developed tools that allow users' to manually assign subjective emotional ratings on visual interfaces [9] or to visualize emotional reactions using biometric storyboards [10].

While these research streams have produced interesting results, they are not easily transferable to new contexts of use, as they are based on internal information from the interactive system (e.g., video game logs, application events, or areas of interest). To address these issues, we developed a new visualization method, in the form of heatmaps, which highlights the areas where users were looking when they experienced specific cognitive and emotional states with a higher frequency, called UX heatmaps [12].

Physiological Heatmaps

To produce physiological heatmaps, different emotional (sadness, happiness, surprise, etc.) and cognitive (cognitive load, stress, etc.) states are first inferred from continuous physiological or behavioral signals. These states are then triangulated with eyetracking data and mapped onto an interface to create heatmaps. In other words, physiological data, for example electrodermal activity and heart rate (HR) are synchronized together, along with eyetracking data. A machine learning model is then used to infer an emotional or cognitive state for each gaze. These are then mapped out onto the interface in the form of heatmaps, which in turn highlight the areas where users tend to emotionally or cognitively react more strongly. Figure 1 illustrates heatmaps generated by participant 01 during our session. On the top interface, a negative valence (red) and positive valence (yellow) heatmaps are shown. The web page below, a cognitive load heatmap is presented.

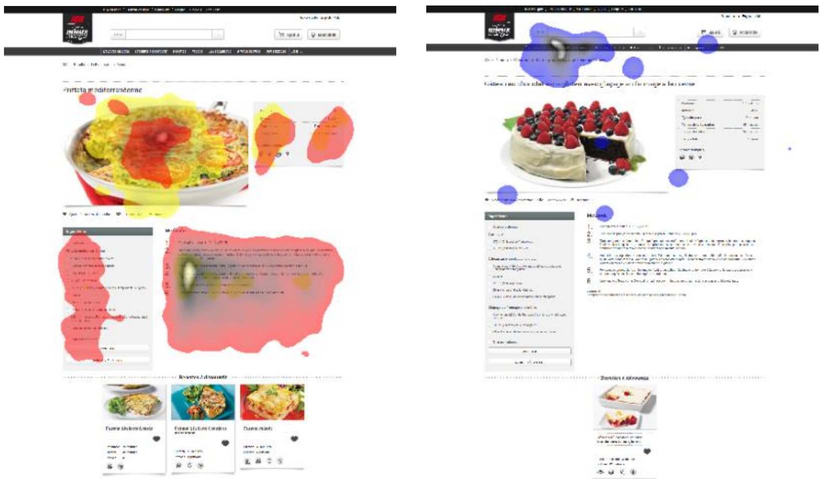


Figure 1 — Negative, positive and cognitive load heatmaps examples. On the left-hand side, negative valence (red) and positive valence (yellow) heatmaps. On the right, a cognitive load heatmap (blue) is illustrated.

Research Method

For this study, a total of 11 UX practitioners and consultants were recruited over a period of 4 weeks. None of the practitioners interviewed had seen our tool prior to the test. Each interview lasted about 1 hour and a half, during which participants were asked to complete a UX evaluation report using the tool following a variation on the think aloud protocol, cooperative evaluation [5]. During the sessions, participants were asked to talk through what they were doing. The interviewer also took on a more active role, by asking questions along the way (e.g. 'why?', what do you think would happen?"). Participants were encouraged to ask for explanations along the way.

Pre-task Interview

We started each session with a preliminary interview to get background information on each participant (see figure 2), such as their number of years of experience in UX as well as their title and main functions within their company, to break the ice and assess their level of qualification. We then gathered their thoughts on physiological measures as a UX evaluation method and assessed their level of familiarity with such methods. Participants had between 2.5 and 24 years of experience in UX, for an average age of 8 years. We interviewed UX directors, consultants, ergonomists and strategists, all of which had heard of physiological measures as an evaluation method in user testing before being approached for this experiment; 7 out of the 11 participants had heard about it while in school, validating the predominance of these methods in academia. Out of all the UX practitioners recruited for this experiment, 8 had previously used physiological measures prior to the study. Eyetracking, being the most popular method overall, was mentioned by all; followed by FaceReader with 3 mentions.

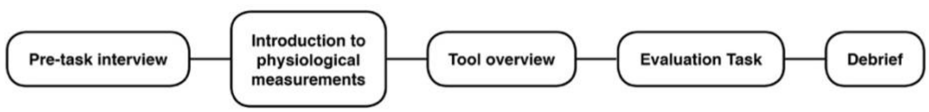


Figure 2 — Experimental procedure

Physiological Measure Introduction and Tutorial

After the introductory discussion, all participants were given a short PowerPoint presentation to introduce them to physiological measures, and were given a tutorial on the tool itself. To do so, we presented each participant with the tool, and went through all the functionalities, buttons and features available to them. We wanted the users to have the same basic knowledge and comprehension of the tool and measures before using it in the completion of a UX evaluation report. The interviewer assisted the participant throughout the experiment, as the goal of the session was not to assess the usability of the tool's interface, but the usefulness of its features and functionalities.

Evaluation Task

During the session, practitioners were asked to complete a user testing evaluation report using our UX heatmaps tool. We therefore provided them with a partially completed PowerPoint report and a 15 participant data set from a previous study. The PowerPoint report included a study summary, a research scenario and qualitative data. We believed this would help UX experts integrate the information on physiological measures quickly and effectively, and also give them a concrete opportunity to use the tool to envision themselves using it in their own practice. First off, participants were briefed on the task at hand, before going through the partially completed report with the interviewer, to put them into context and get a sense of what was required of them. Participants had to complete a total of 2 PowerPoint slides. They were asked to : 1) generate and select data visualizations to include in their report using our tool, 2) interpret the results and 3) provide recommendations to the client. The remainder of the time was used to discuss the advantages and disadvantages of physiological measures as an evaluation method, as well as the tool itself.

Results

Participants made interesting comments regarding physiological measures and our tool, which we will address in the following section. We are only reporting comments made by 3 or more participants. Interviewees mentioned the following as the ways in which they would use our tool in their own practice:

- Provide new avenues for research
- Form and confirm research hypotheses
- Guide discussions during interviews
- Confirm and validate findings
- Elaborate evaluation tests

The main contribution of our tool, as stated by 5 participants, is the comparison and the juxtaposition of different emotional and cognitive states. As participant 07 explained, "there are simply no other tools available that make this essential data accessible to us". Participants also mentioned the collaborative potential of our tool. The visualizations generated could be used communicate information to the various members of the design team, as well as with clients and management. For example, participant 10 suggested that the visualization generated could be shared with designers for them "to better understand the impact of their creative freedoms on the user".

Data Contextualization and Interpretation

Our goal in creating our tool was to address one of the main concerns associated with the use of physiological measures, the interpretation of physiological and behavioral signals. We set out to do these interviews with industry practitioners to find out how we fared at the task. Overall, participants found physiological heatmaps easy to interpret. As six participants mentioned, the visualizations were clear, intuitive and wielded powerful results that facilitated the interpretation of physiological signals.

Users stated that our tool was also easy to understand from a client’s perspective. For example, participant 08 felt that customers would appreciate seeing the emotions generated by problematic areas directly onto the interface, adding "it goes beyond qualitative insight". Two participants found the interpretation of the data to be difficult without prior knowledge of physiological measures, one practitioner adding "the learning curve is relatively mild; the analysis should become more natural with time".

As illustrated in figure 3, participants were able to make insightful and actionable recommendations based on the visualizations generated with our tool; on the left, a gaze (green), a positive (yellow) and negative valence (red) heatmaps generated by P04. Although the focal element of the page was the text area below, the image clearly elicited positive emotions, while negative emotions or displeasure was experienced by users in correlation to the instructions of the recipe. By comparing regions of negative and positive valence, the practitioner identified problematic areas of the interface and was able to highlight the graphical elements behind them. Based on these results, the practitioner recommended to increase positive emotions and arousal experienced on the page by adding visual elements, such as videos and pictures, and revising the presentation of the recipe's instructions to avoid superfluous text areas.

When asked about their intent to reuse the tool, 10 out of the 11 practitioners interviewed stated that they would use the tool in their practice. However, when inquired further, 6 of them declared that their use of the tool would depend on the projects, using it only in the assignments where emotions are an important component or if clients specifically requested them to use physiological signals.

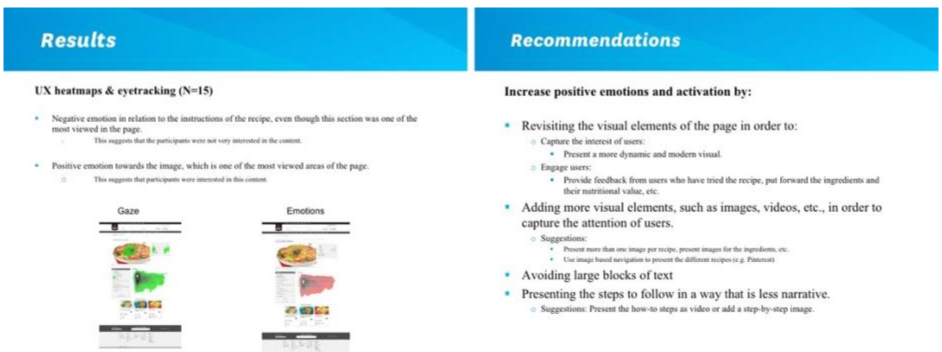


Figure 3 — An example of a completed report by a participant, translated from French to English.

Discussion

When developing new UX evaluation tools using physiological measures, the ability to locate issues, the ease of use and interpretation and the reduction of analysis time represent important factors. Overall, participants found that physiological signals would be integrated more easily into their practice using our tool. Participants suggested the following improvements to UX heatmaps to further facilitate the adoption of physiological measures their current practice:

- The addition of an event timeline, or replay feature, to better understand overlapping UX heatmaps, to see the order in which the different emotional and cognitive states occurred. This would help with the interpretation of the visualizations.

- The inclusion of supplementary information, collected from traditional UX methods, such as participants' profiles and usability metrics. This would help them to integrate physiological methodologies more easily to the methods they currently use in their practice.

- The automatization of certain functions, such as groups and layer creation, to accelerate the interpretation of the visualizations generated with our tool. This would help them fit this analysis within their short development cycles.

Although our tool makes physiological measures more accessible to UX practitioners by addressing the interpretation of signals, there remains a lot of work to be done regarding some of the more technical aspects of physiological measurements. Participants expressed concerns regarding the time constraints pertaining to the actual experimental setup of such user testing, for example the selection of signals and the placement of sensors, as well as the resources needed to run the experiment. Knowledge of physiological measures is still needed, as the signals used for physiological heatmaps should be selected according to the psychological variables of interest (e.g. emotion, cognitive load, etc.). Physiological measurements still represent important time and financial constraints, as data collection, experimental setup and data extraction still have to be overseen by the UX professional.

As mentioned above, practitioners who use physiological measures are doing so in particular projects only, i.e. projects that require the evaluation of emotions or if these measures are requested by the client. This translates into a steep and ever present learning curve, as practitioners must re-learn how to use the tools and materials associated with the data collection of physiological signals at each use. Therefore, the practitioners are never able to develop an expertise. Unable to justify the financial investment due to sparse usage of such tools, practitioners often end up renting the equipment, which is very costly.

Having practitioners use our UX heatmaps tool in the completion of an actual user testing evaluation report following a cooperative evaluation protocol yielded great results. We would recommend using this method for the evaluation of new tools and methodologies as:

- Participants felt comfortable to criticize physiological methodologies and our tool
- Provided a more relaxed atmosphere where participants could see themselves as collaborators rather than as experimental subjects
- Helped them take ownership of the tool and explore the functionalities it offered
- Helped us get insights as to how this tool would be received in the community.

We had hoped that the interview process would generate new ideas and avenues of research, in addition to potential improvements to our tool. However, this did not occur. We may have had more in-depth insights as to new functionalities had we :

- Interviewed practitioners who were more familiar with or used physiological measures in their current practice
- Had practitioners used the tool over longer periods of time. In the sessions, interviewees had only between 25 to 35 minutes to use the tool and complete their task.

Conclusion

The use of physiological measures, in combination with traditional methods, could help practitioners to better measure UX, as they each provide complementary information on how users feel about a system, game, or web interface. [6] While traditional evaluation methods can offer episodic data, i.e. before or after the interaction, physiological measures can provide moment-to-moment information [13]. The addition of physiological measures can help us identify the cognitive and emotional reactions users experienced using an interface, while a post-task interview can help us delve further, after we have identified these emotions.

The main research and development activities we undertake at the Tech3Lab aim at facilitating and fostering the adoption of new methodologies, such as eyetracking and physiological measures, in the fields of UX design and research. A first step towards this direction was the development of a physiological heatmaps tool to allow simpler and richer interpretation of physiological signals for UI evaluation. The interviews we conducted with UX practitioners were very helpful, in that they provided guidelines and user requirements insights for us to use in the development of future iterations to facilitate furthermore the adoption of physiological methodologies. Our next step will be to continue to develop our functionalities as well as find ways to simplify the data processing sequence associated with physiological signals, working closely with ergonomists and consultants of the industry to do so.

Acknowledgments

Authors want to thank Brendan Scully for manuscript revision and the UX practitioners who participated to this study. This research was supported by NSERC (National Sciences and Engineering Research Council of Canada).

References

1. Hassenzahl, M., Tractinsky, N. : User experience-a research agenda. Behav. Inf. Technol.25 (2), 91–97 (2006).

2. Zhihong, Z., Pantic, M., Roisman, G.I., Huang, T.S. : A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Trans. Pattern Anal. Mach. Intell. 31(1), 39–58 (2009).

3. Calvo, R.A., D'Mello, S. : Affect detection: an interdisciplinary review of models, methods, and their applications. IEEE Trans. Affect. Comput. 1(1), 18–37 (2010).

4. Georges, V., Courtemanche, F., Sénécal, S., Baccino, T., Fredette, M., Léger, P.M. : (Forthcoming). UX heatmaps: mapping user experience on visual interfaces. In : 34rd Annual ACM Conference on Human Factors in Computing Systems.

5. Dix, A., et al. : Human-computer interaction. Pearson Education (2004).

6. Roto, V., Vermeeren, A.P.O.S., Väänänen-Vainio-Mattila, K., Law, E., Obrist, M. : Course note : user experience evaluation methods - which method to choose? In : Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (Paris, France). ACM (2013).

7. Zaman, B., Shrimpton-Smith, T. : The FaceReader: measuring instant fun of use. In : Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, pp. 457–460. ACM, Oslo (2006).

8. Boucsein, W.: Electrodermal Activity. Springer, Berlin (2012).

9. Huisman, G., van Hout, M., van Dijk, E., van der Geest, T., Heylen, D. : LEMtool: measuring emotions in visual interfaces. In : Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 351–360. ACM, Paris (2013).

10. Mirza-Babaei, P., Nacke, L.E., Gregory, J., Collins, N., Fitzpatrick, G. : How does it play better?: exploring user testing and biometric storyboards in games user research.

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2013), pp. 1499–1508. ACM, New York (2013).

11. Hassenzahl, M., Burmester, M., Koller, F. : AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualitat (AttracDiff: A questionnaire to measure perceived hedonic and pragmatic quality). In : Ziegler, J., Szwillus, G. (eds.), Mensch & Computer 2003. Interaktion in Bewegung, pp. 187–196. B.G. Teubner, Stuttgart, Leipzig (2003).

12. Courtemanche, F., Léger, P.-M., Frédette, M., Sénécal, S., Georges, V., Dufresne, A. : Method and Product for Visualizing the Emotions of a User, Provisional patent application. US 62/121, 552, p. 14 (2015).

13. Virpi, R., et al. : All about UX (2012). http://www.allaboutux.org/.

14. Riedl, R., Léger, P.-M. : Fundamentals of NeuroIS. Springer, Berlin (2016).

15. Kivikangas, M., Nacke, L., Ravaja, N. : Developing a triangulation system for digital game events, observational video, and psychophysiological data to study emotional responses to a virtual character. Entertainment Comput. 2(1), 11–16 (2011).

Conclusion

3. Sommaire

L'objectif principal de notre outil de visualisation est de faciliter l'interprétation des signaux neurophysiologiques, afin de réduire les obstacles associés à l'utilisation de ces mesures dans le domaine de l'expérience utilisateur. Chacun des trois articles de ce mémoire détaille une des phases du projet de recherche, soit l'idéation, la conception et l'évaluation de l'outil.

Dans le premier article, nous proposons une nouvelle méthode d'évaluation de l'état émotionnel et cognitif des utilisateurs basée sur les signaux neurophysiologiques de ceux-ci. Le but recherché par la méthode étant d'informer les praticiens et les chercheurs du domaine de l'UX sur la manière dont les états émotionnels et cognitifs ressentis par les utilisateurs durant l'interaction humain-machines affectent leurs comportements et leurs motivations. Le deuxième article présente l'outil d'évaluation, dont l'élaboration nous a permis de relever les trois défis suivants concernant l'utilisation de mesures physiologiques dans le prototypage et l'évaluation d'interfaces utilisateur par des experts du domaine de l'UX : 1) la synchronisation de signaux multiples provenant de différents appareils, 2) l'inférence de construits psychophysiologiques multiples et 3) la contextualisation des données à l'aide de cartes de chaleur. Le troisième article de ce mémoire met en évidence les commentaires recueillis auprès de praticiens de l'UX sur les manières de faciliter l'adoption des mesures physiologiques dans leur pratique. Plus amples exemples de cartes de chaleur physiologiques et de leur interprétation par des praticiens se trouvent en annexe.

3.1. Implications pour la recherche

En recherche, les besoins d'évaluation de l'expérience utilisateur résident dans la validation et la compréhension d'évènements ou phénomènes, sur la base d'hypothèses [10]. Du point de vue des chercheurs, un outil comme celui proposé dans ce mémoire peut aussi informer les chercheurs sur la relation entre les états émotionnels et cognitifs expérimentés par les utilisateurs et les comportements de ceux-ci, par exemple les activités d'échanges et de transactions effectuées par l'intermédiaire d'Internet.

3.2. Implications pour l'industrie

L'évaluation de l'expérience utilisateur en industrie repose sur l'analyse et la communication adéquate de résultats, afin d'améliorer l'UX [10]. L'outil peut donc aider les professionnels du domaine en facilitant la prise de décision et le partage des résultats, en plus de démontrer le potentiel des mesures neurophysiologiques en UX.

Pour les praticiens, ce type de visualisation peut aider à prendre des décisions plus éclairées dans des contextes de comparaison de prototypes ou de tests A/B, en fournissant des données quantitatives sur lesquelles baser la prise de décision. De plus, l'outil facilite l'identification des zones problématiques d'une interface pouvant affecter l'utilisateur dans l'accomplissement d'une tâche ou qui pourraient servir d'irritant dans l'utilisation d'un produit, d'un système ou d'un service. Par exemple, dans le contexte d'une analyse comparative de plusieurs prototypes ou une analyse exploratoire, les cartes de chaleur permettent de contextualiser les états émotionnels et construits cognitifs ressentis par les utilisateurs durant leurs interactions en mettant en évidence les éléments graphiques ayant causé ces états.

De plus, notre outil de visualisation fournit aux praticiens un moyen de communication et de collaboration que tous peuvent utiliser et comprendre, afin de faciliter la collaboration entre professionnels ayant des parcours différents.

Les résultats de ces articles démontrent le potentiel des mesures neurophysiologiques en UX pour mesurer l'expérience utilisateur. Les mesures physiologiques, en combinaison avec les méthodes d'évaluation traditionnelles, peuvent donc aider les praticiens à mieux évaluer l'expérience utilisateur, puisqu'ils fournissent chacune des informations complémentaires sur la façon dont les émotions, les perceptions et les réactions des utilisateurs affectent ceux-ci.

4. Limites et recherche future

L’utilisation de l’outil de visualisation présente des défis supplémentaires liés à l’utilisation des mesures neurophysiologiques, tels que la manipulation d’équipement de collecte et l’utilisation de systèmes d’acquisition de données provenant de sources externes. En d’autres termes, bien que notre outil de visualisation rende les mesures neurophysiologiques plus accessibles aux praticiens du domaine de l’expérience utilisateur en facilitant l’interprétation des signaux, certains aspects plus techniques associés aux mesures neurophysiologiques restent problématiques. Le choix des mesures neurophysiologiques, la manipulation des capteurs, la collecte de données, la configuration expérimentale et l’extraction de données doivent toujours être entrepris par les praticiens et chercheurs, ce qui représente d’importantes contraintes temporelles et financières. De plus, la contrainte associée à l’expertise spécifique demeure, la mise en place de la collecte requérant des connaissances expertes dans diverses. De la recherche supplémentaire sur la diminution de ces obstacles dans les domaines de la conception et de la recherche de l’expérience utilisateur pourrait donc être pertinent.

Bibliographie

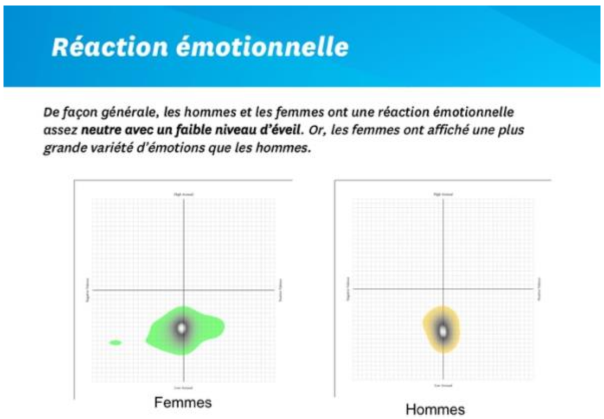
1. Hassenzahl, M. and N. Tractinsky, *User experience-a research agenda*. Behaviour & information technology, 2006. 25(2): p. 91-97.
2. Whiteside, J. and D. Wixon, *The dialectic of usability engineering*. Human-Computer Interac-tion—INTERACT, 1987. 87: p. 17-20.
3. Février, F., et al., *La prise en compte des affects dans le domaine des interactions homme-machine: quels modèles, quelles méthodes, quels bénéfices?* Le travail humain, 2011. 74(2): p. 183-201.
4. Hassenzahl, M., *The thing and I: understanding the relationship between user and product*, in *Funology*. 2003, Springer. p. 31-42.
5. Larousse, P., *Nouveau petit Larousse: dictionnaire encyclopédique pour tous*. 1969: Librairie Larousse.
6. Garrett, J.J., *Elements of user experience, the: user-centered design for the web and beyond*. 2010: Pearson Education.
7. Jallat, F., et al., *Gestion de la relation client: Total relationship management, Big data et Marketing mobile*. 2014, Paris Dauphine University.
8. Bevan, N. *What is the difference between the purpose of usability and user experience evaluation methods*. in *Proceedings of the Workshop UXEM*. 2009.
9. Robert, J.-M. and A. Lesage, *From usability to user experience with interactive systems*. Handbook of Human-machine interaction, 2010: p. 303-332.
10. Virpi Roto, A.P.V., K Väänänen-Vainio-Mattila, Effie LC. Law and Marianna Obrist. *User Experience Evaluation Methods - Which Method to Choose?* in *SIGCHI Conference on Human Factors in Computing Systems*. 2013. Paris, France: ACM.
11. Lallemand, C. and G. Gronier, *Méthodes de design UX*. Paris, Eyrolles, 2016.
12. Dix, A., et al., *Human-computer interaction: Pearson Education*. 2004.
13. van den Broek, E.L., et al., *Prerequisites for affective signal processing (ASP)-Part IV*. 2010.
14. de Guinea, A.O., R. Titah, and P.-M. Léger, *Explicit and implicit antecedents of users' behavioral beliefs in information systems: A neuropsychological investigation*. Journal of Management Information Systems, 2014. 30(4): p. 179-210.
15. Robinson, M.D. and G.L. Clore, *Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes*. Journal of personality and social psychology, 2002. 83(1): p. 198.
16. Zeng, Z., et al., *A survey of affect recognition methods: Audio, visual, and spontaneous expressions*. IEEE transactions on pattern analysis and machine intelligence, 2009. 31(1): p. 39-58.
17. Calvo, R.A. and S. D'Mello, *Affect detection: An interdisciplinary review of models, methods, and their applications*. IEEE Transactions on affective computing, 2010. 1(1): p. 18-37.
18. Ganglbauer, E., et al. *Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility*. in *Workshop on user experience evaluation methods in product development*. 2009.

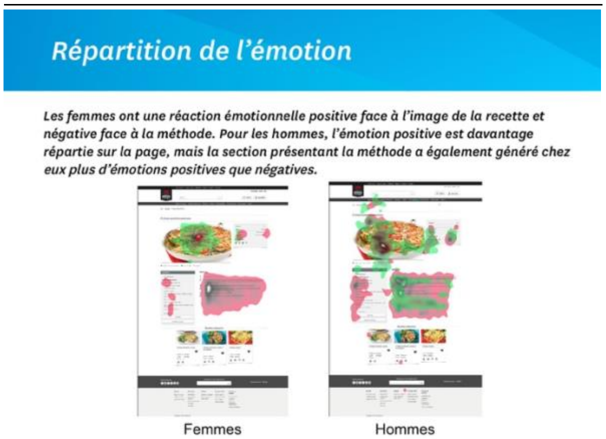
19. Haapalainen, E., et al. *Psycho-physiological measures for assessing cognitive load*. in *Proceedings of the 12th ACM international conference on Ubiquitous computing*. 2010. ACM.
20. Dawson, M.E., A.M. Schell, and D.L. Filion, *The electrodermal system.* Handbook of psychophysiology, 2007. 2: p. 200-223.
21. Boucsein, W., *Electrodermal activity*. 2012: Springer Science & Business Media.
22. Shi, Y., et al. *Galvanic skin response (GSR) as an index of cognitive load*. in *CHI'07 extended abstracts on Human factors in computing systems*. 2007. ACM.
23. Goldwater, B.C., *Psychological significance of pupillary movements.* Psychological bulletin, 1972. 77(5): p. 340.
24. Karatekin, C., J.W. Couperus, and D.J. Marcus, *Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses.* Psychophysiology, 2004. 41(2): p. 175-185.
25. Piquado, T., D. Isaacowitz, and A. Wingfield, *Pupillometry as a measure of cognitive effort in younger and older adults.* Psychophysiology, 2010. 47(3): p. 560-569.
26. Nielsen, J. and K. Pernice, *Eyetracking web usability*. 2010: New Riders.
27. Holmqvist, K., et al., *Eye tracking: A comprehensive guide to methods and measures*. 2011: OUP Oxford.
28. Calvo, R.A., et al., *The Oxford handbook of affective computing*. 2015: Oxford Library of Psychology.
29. Zaman, B. and T. Shrimpton-Smith. *The FaceReader: Measuring instant fun of use*. in *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*. 2006. ACM.
30. McDuff, D., et al. *Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2013.
31. Ekman, P. and W.V. Friesen, *Facial action coding system.* 1977.
32. Ward, R.D. and P.H. Marsden, *Physiological responses to different WEB page designs.* International Journal of Human-Computer Studies, 2003. 59(1): p. 199-212.
33. King, M.F. and G.C. Bruner, *Social desirability bias: A neglected aspect of validity testing.* Psychology and Marketing, 2000. 17(2): p. 79-103.

Annexe

Des exemples de cartes de chaleur physiologiques et de leur interprétation par des praticiens récoltés lors de la troisième étude de ce mémoire.

<u>Participant 09</u>





Répartition de l'attention

Pour les hommes et les femmes, la zone regardée le plus longtemps correspond à la méthode. On remarque que les hommes ont porté peu attention à l'image par rapport aux femmes.

Femmes

Hommes

Principaux constats

- ***Globalement, les femmes portent davantage attention aux éléments visuels (image dans ce cas-ci), qui génère de l'émotion positive chez elle. Les hommes parcourent plutôt la page dans son ensemble.***
- ***La méthode, qui est une section constituée uniquement de texte, génère de l'émotion négative chez les femmes et les hommes bien qu'il s'agisse de l'élément qui attire le plus leur attention.***
  - ***Ces résultats ne sont pas surprenants considérant que la plupart des participants ont trouvé que l'interface n'était pas suffisamment visuelle et ont eu du mal à repérer certaines informations clés comme les valeurs nutritives ou le niveau de difficulté de la recette.***

Recommandations

Rendre la présentation du contenu plus visuel afin de générer plus d'émotion positive chez les utilisateurs – particulièrement les femmes et rendre certains éléments clés plus faciles à repérer sur l'interface.

1-Utiliser un schéma pour **présenter la méthode de façon plus dynamique** ou, si possible, utiliser des images illustrant le processus.

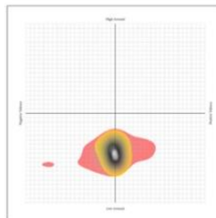
2-**Faire ressortir les informations nutritionnelles davantage** en utilisant une couleur contrastante avec le reste de la page pour qu'elles soient plus faciles à repérer sur la page. Actuellement, cet encadré est peu regardé.

Participant 07

Résultats

Compléter en utilisant des résultats provenant du logiciel testé.

Les femmes ont une réaction plus émotionnelle lors de la consultation de recettes. Cette réaction peut être positive ou négative.



Résultats

Compléter en utilisant des résultats provenant du logiciel testé.

Les gens qui cherchent une recette en particulier vont avoir une interaction bcp plus superficielle (moyenne 3 sec) que ceux qui consultent une recette pour la faire (moyenne 2 min).

Recommandations

Émettre 2 recommandations.

- ***Adapter le contenu des infolettres en fonction du sexe du destinataire.***
  - ***Femmes: miser sur l'émotion dans le vocabulaire.***
  - ***Hommes: plus sur les aspects fonctionnels de la recette (prix, temps de préparation, etc.)***
- ***Popup après 2 minutes qui suggère d'ajouter les ingrédients à la liste d'épicerie ou met en valeur un ingrédient présentement en spécial.***
- ***Si la personne est connecté à son compte et elle reste plus de 2 minutes devant une recette, lui resuggérer la recette dans une prochaine infolettre.***
