



HAL
open science

Using user emotions to trigger UI adaptation

Julián Andrés Galindo, Sophie Dupuy-Chessa, Nadine Mandran, Éric Céret

► **To cite this version:**

Julián Andrés Galindo, Sophie Dupuy-Chessa, Nadine Mandran, Éric Céret. Using user emotions to trigger UI adaptation. 2018. hal-01824406

HAL Id: hal-01824406

<https://hal.science/hal-01824406v1>

Preprint submitted on 27 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using user emotions to trigger UI adaptation

Julián Andrés Galindo
Univ. Grenoble Alpes, CNRS, LIG
F-38000, Grenoble, France

Escuela Politécnica Nacional, Department of Informatics
and Computer Science, Quito, Ecuador
julian.galindo@epn.edu.ec

Sophie Dupuy-Chessa, Nadine Mandran, Eric Céret
Univ. Grenoble Alpes, CNRS, LIG
F-38000, Grenoble, France

Firstname.Name@univ-grenoble-alpes.fr

Abstract—User experience is an important user interface quality factor that needs to be maintained during interaction. With this goal, we propose to adapt user interfaces to users' emotions. This requires being able to detect at runtime negative emotions to trigger adaptations. However, emotions depend on users' gender and age. This makes it difficult to identify negative emotions and their cause. This paper studies the categorization of emotions depending on users' age and gender and two user experience dimensions (aesthetics and usability). Thanks to an experiment, it proposes some emotions thresholds that can be used for detecting usability and aesthetics problems from users' age and gender. These results are used in a user interface adaptation system to infer negative emotions and user interface problems.

Keywords— *UI adaptation, emotion, usability, aesthetics*

I. INTRODUCTION

Beyond usability, user experience (UX) is nowadays recognized as an important quality factor to make systems or software successful in terms of user take-up and frequency of usage [1]. For instance, according to [2], UX depends on dimensions like emotion, aesthetics or visual appearance, identification, stimulation, meaning/value or even fun, enjoyment, pleasure, or flow. Among these dimensions, the importance of usability and aesthetics is recognized [3]–[5]. So, both of them need to be considered while designing user interfaces (UI) [6], [7].

Beyond design, it would be interesting to check UX at runtime and improve it if necessary. To achieve a good UI quality in any context of use (i.e. user, platform and environment), plasticity [8] proposes to adapt UI to the context while preserving user-centred properties. In a similar way, our goal is to preserve or improve UX at runtime, by proposing UI adaptations. Adaptations can concern aesthetics or usability. They can be triggered by the detection of negative emotions [9], as these emotions can express a problem with the UI.

Complexity is then related to the categorization of emotions as negative because they depend on various individual

characteristics such as age [10], [11] and gender [11]–[13]. It is even higher while considering categorization at runtime.

We present an experiment that allows us to categorize users' emotions according to the levels of UI usability and/or aesthetics and to the users' age and gender. Indeed, even if there are other emotion's sources (e.g. personality, day time-line, weather or social activities) [14], we focus on these two static sources for stating the interest of the approach.

From these categorized users' emotions, it is then possible to identify if a problem exists in UI usability and/or aesthetics and to choose if some adaptation must be triggered. Lastly, these results are used to trigger adaptations in the prototype proposed in [14].

Our long-term motivation is to propose a system able to adapt UI to emotions. The contribution of this work is a step toward this goal by making it possible to identify usability or aesthetics problems thanks to emotions.

The next section of the paper presents work related to affective computing and adaptation to emotions. Section 3 describes the experiment and its results. In section 3, we use these results to propose some triggers for UI adaptation. Section 4 presents a prototype to implement them. The paper ends with some conclusions and future work.

II. RELEATED WORK

In human-computer interaction, users' emotions are recognized as being important in the design and evaluation processes [15]. They cannot be ignored while designing for UX [15]. In this context, the field of Affective computing [16] aims at creating « machines that relate to, arise from, or deliberately influence emotion or other affective phenomena » according to [15]. It provides models [17], [18] and tools such as FaceReader¹ or Affdex² to measure affects in terms of basic emotions from Ekman's model (surprise, happiness, sadness, disgust, anger, fear) [17] plus neutral. It also studies emotions related to users' features such as learnability, performance, communication in some specific domains such as tutoring systems [19]–[21].

Beyond models, emotions can be used at runtime in closed-loop applications [22]: emotions are measured and interpreted in terms of affects; then, based on the affects, a decision is made

¹ www.noldus.com/human-behavior-research/products/facereader

² www.affectiva.com/solutions/affdex/

concerning the applicable actions; these actions are executed; and the loop starts again with emotions measures. In this way, adaptations can be triggered either on the functional part of the application [23] – e.g. for changing the content – or on the UI part – e.g. for adapting the design or the interaction modality. In our work, we focus on the UI adaptation.

Only few works [24]–[27] studies UI adaptation based on emotions even if the user’s emotional state is cited as one of the elements of the context of use for UI adaptation [9].

[24] proposes an approach which aims at identifying the cause of an emotion from the history of interactions and then at proposing an appropriate adaptation. In this work, interaction concerns humans and robots, which is a very specific kind of human-computer interaction domain, particularly in terms of usability and aesthetics. Moreover, the approach does not seem to have been implemented in systems.

The ABAIS approach (Affect and Belief Adaptive Interface System) [25] follows user’s anxiety while interacting with a complex air force system. In this work, adaptation is triggered by only one emotion (stress) and pilot’s beliefs (e.g. hostile aircraft approach) in a specific context. The results cannot be reused in other contexts for more widespread UIs. Interestingly, the proposed adaptations can affect icons, displays, notifications and custom configuration which reflects both usability and aesthetics changes. But the distinction between usability and aesthetics problems is not realized as the goal is not to improve UX.

Nasoz [26] proposed a more general approach for UI adaptation to emotions. It consists in implementing an adaptive virtual reality system relying on the recognition of affective states from physiological signals. It includes a user model with the following features: age, gender, recognized emotion (sadness, anger, surprise, fear, frustration, and amusement) and frequency of experiencing a specific emotion. From these data, an avatar mirroring the user can be shown and a Bayesian belief network can propose some actions adapted to user’s emotions, like “make a joke” or “suggest to stop the car”. Here a personalized avatar (skin colours, voices, hair, make-up, accessories) and a background are proposed based on user’s preferences. Then adaptation is the mirroring of the user’s current emotional state with the avatar design (face shape and orientation changes). The detected emotions are not related to the UI but to the driving situation. So, if the work proves that it is possible to trigger adaptations from several emotions, it does not explicit how detecting UI usability or aesthetics problems.

Finally, [27] proposes an architecture where automatically detected emotions can trigger UI adaptation based on patterns. Adaptation can be related to colours or structure and is performed at runtime. The approach is very interesting as it proposes an adaptation of UI aesthetics and usability based on emotions at runtime. It has very similar goals to ours. However, adaptations are based on one dominant emotion over a period of time. This emotion is compared with a dominant emotion determined at design time from a panel of users for each application. The way this comparison is realized is not made precise: the period of time and the threshold that triggers

adaptation are not specified. There is also no evidence about how to identify the cause of non-expected emotions and then proposing appropriate modifications.

To sum up, only few works studied UI adaptations based on emotions at runtime. None of them seems able to detect at runtime the cause of a UX problem in terms of usability or aesthetics and then to appropriately trigger UI adaptations.

The next section presents an experiment, which allows us to identify usability and aesthetics problems based on age and gender at runtime. Then these results can be integrated in a prototype for UI adaptation proposed in section 4.

III. EXPERIMENTAL STUDY

One of the main challenges of UI adaptation to emotions is to identify, at runtime, emotions that must trigger adaptations. To address it, we conduct an experimental study, which allows us to better understand users’ emotions based on their age and gender and to analyze it in regard with the level of usability and aesthetics. We take advantages of the seven emotions (anger, contempt, disgust, sadness, fear, neutral, happiness) detected from captured images by commercial tools like FaceReader in order to obtain a fine grain analysis and to detect negative emotions at runtime.

A. Goal and hypothesis

The experiment aims at identifying usability and/or aesthetics problems based on users’ age and gender at runtime. To achieve this goal, it is necessary to detect fine grain emotions at runtime. From them, some categories of emotions, depending on users’ profile as well as UI aesthetics and usability quality levels, can be identified. They represent some typical emotional behaviours, defined by emotions intervals. These intervals determine emotions thresholds from which it is possible to detect a problem. Our hypotheses are thus:

H1: it is possible to categorize emotions depending on users’ age and gender and on UI aesthetics and usability levels;

H2: there exist emotions thresholds for detecting usability and aesthetics problems from users’ age and gender.

B. Experimental protocol

Four UI versions of a travel website were designed and developed [28]. The websites vary by two factors or UI qualities: usability and aesthetics. Consequently, the combination of two values (low vs. high) for these two factors (aesthetics and usability) released four versions of the website as evidenced in Fig. 1. [28] shows the relevance of using such websites versions to identify the impact of usability and aesthetics on UX.

These four used websites followed (1) usability and (2) aesthetics design modifications previously implemented by [28]. First, usability was considered at the task level (e.g. interaction distance and workload) as well as at concrete user interface (CUI) level. The CUI was redesigned for the low usable version for increasing the workload by changing the UI components or the navigation easiness (e.g. increasing



Figure 1. Four versions of a web site user interface.

interaction distance or adding needless articulatory subtasks). Second, the aesthetics quality level (high vs. low) was achieved by modifications only in the graphic factors. As illustrated in Fig. 1 (e.g. website version number 1 vs number 2), this included the background-colour (bright vs. light), the framework shape (square vs. rounded), and the format style (non-aligned vs. aligned).

Each participant was assigned to one version of the travel website. While he/she was interacting with the website, his/her emotions were detected automatically based on captured user's photos in a time period of 10 seconds. Photos are sent to an emotion detection tool, Microsoft Emotions API³. This tool detects five negative (anger, contempt, disgust, sadness, fear) and one positive (happiness) emotions which were stated as universal by Ekman [17]. It also detects neutral even if it is not part of the set of universal emotions, it is important to differentiate a calm (neutral) against an arousal user response (six Ekman emotions). This set of 7 emotions is recognized as a basic set proposed by many emotion detection tools based on images. Secondly, to ensure that participants are involved enough in the web interaction, they had to answer an online questionnaire after each task.

Three tasks are required for each participant: (1) look for a 5-stars hotel which offers a visit in the desert, (2) look for a specific hotel with must include a breakfast plan and (3) look for a hotel circuit in a specific given location.

C. Participants and interaction data

45 persons participated to the study, with an average of approximately 11 participants per website and a distribution of: 13 (website1 - low aesthetics/high usability), 10 (website2 - high aesthetics/high usability), 12 (website3 - low aesthetics/low usability), and 10 (website4 - high aesthetics/low usability). The experiment sample included 53% (24) male from 19 to 67 years old and 47% (21) female from 23 to 63 years old participants (mean male age 34, $\sigma=11.66$ and mean female age 27 with $\sigma=12.25$).

For each participant, we registered the UI version used and his/her seven emotions detected every 10 seconds as an experimental time period. This period of time was chosen because a face macro-expression takes almost 4 seconds [29] and it is thus expected that a 10-second-period is good enough to capture a proper emotional response. As some images are not good enough for being useable by the emotion detection tool, we finally obtain 1730 usable images with an average of 38 images by user.

Moreover, each image is marked with a timestamp, the UI in use (website version and current page) and the user's characteristics. We name an observation the tuple of data with the timestamp, the seven detected emotions (disgust, anger, fear, sadness, contempt, neutral and happiness), the current UI and the user's age and gender. An observation corresponds to

³ www.microsoft.com/cognitive-services/en-us/emotion-api

an image, so we got 1730 observations with 11 data inside each tuple, giving a total of 19,030 data.

D. Data analysis

This section discusses the intermediary results while the next section presents results related to the two experimental hypotheses.

The data analysis process, illustrated in figure 2, started with a multivariable analysis ❶ called Principal Component Analysis (PCA) in order to find discrimination values in socio-demographic variables. Principal Components Analysis is a statistical analysis which extracts some significant factors that resume data information.

It uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called factors or principal components. Only significant factors are kept (eigenvalues greater than 1). Here PCA allowed us to observe a difference between men and women and between ages in emotions expression in accordance with previous studies [10]–[13].

Then we divided the sample into 4 categories depending on gender and age ❷. For age, we aimed at keeping the same number of persons in each age category in order to have enough individuals in each category to conduce a statistical analysis. The threshold was 27 years old. So, the 4 categories are: women less than 27, women more than 27, men less than 27 and men more than 27.

From these categories, we identified similar emotion behaviours thanks to another PCA ❸ and a Hierarchical Cluster Analysis (HCA) ❹. A PCA analysis was realized to study the links between emotions by individual. It limits the variance quantity between individuals and removes data noise. PCA creates new coordinates (scores) of emotions for each observation.

Table 1 shows how two retrieved principal components (PC1 &

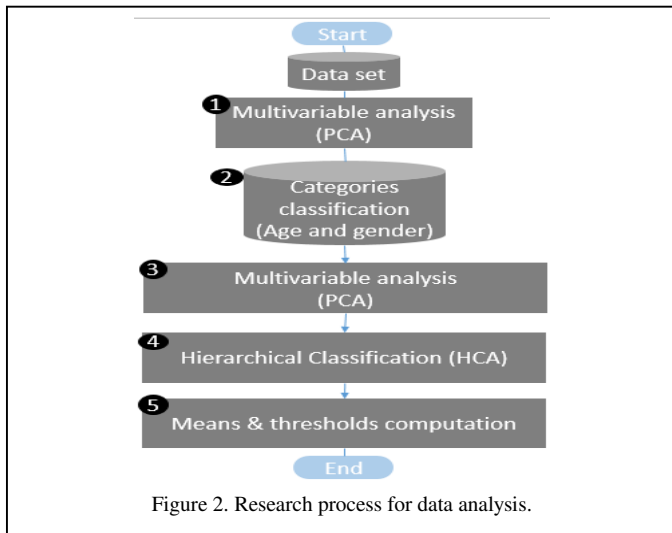


Figure 2. Research process for data analysis.

PC2) in each case (women < 27, women >= 27, men < 27, men

	women < 27		women >= 27	
	σ_x	σ^2	σ_x	σ^2
PC1	1,66	0,40	1,58	0,36
PC2	1,33	0,25	1,28	0,23
PC3	0,92	0,12	1,05	0,16
PC4	0,86	0,11	0,84	0,10
PC5	0,63	0,06	0,76	0,08
PC6	0,49	0,03	0,56	0,05
PC7	0,47	0,03	0,41	0,02
	men < 27		men >= 27	
	σ_x	σ^2	σ_x	σ^2
PC1	1,96	0,55	1,93	0,53
PC2	1,07	0,16	1,09	0,17
PC3	0,85	0,10	0,96	0,13
PC4	0,79	0,09	0,79	0,09
PC5	0,67	0,06	0,58	0,05
PC6	0,39	0,02	0,33	0,02
PC7	0,24	0,01	0,31	0,01

σ_x : standard deviation and σ^2 : variance

Table 1: Standard deviation and variance values for each principal component across women and men clusters.

>= 27) contributes more than others based on the variance(σ^2). For instance, PC1 and PC2 reached in total 65% of variance while all other ones accumulated 35% in women less than 27. Similarly, the two components (PC1 & PC2) accumulated more variance in the three remain cases with 59, 71 and 70 per cent respectively.

	Cluster 1	Cluster 2	Cluster 3
Women < 27	216	259	304
Women > 27	74	508	28
Men < 27	337	558	310
Men > 27	422	259	97

Table 2 –Number of scores in the HCA process

Furthermore, except from women over 27 years old, the two mentioned principal components (PC1 & PC2) were the most important ones by releasing standard deviation values (σ_x) higher than 1 for each case.

These new coordinates were computed with a Hierarchical Cluster Analysis ❹. The goal of this technique is to create some clusters sufficiently distinct from each other but with an intrinsic homogeneity. Here the Hierarchical Cluster Analysis produced significant partitions with 3 clusters for each case (women<27, women>27, men>27, men<27, see Table 2).

Then for each cluster, we studied the relationship between emotions detected and the website version used (low/high

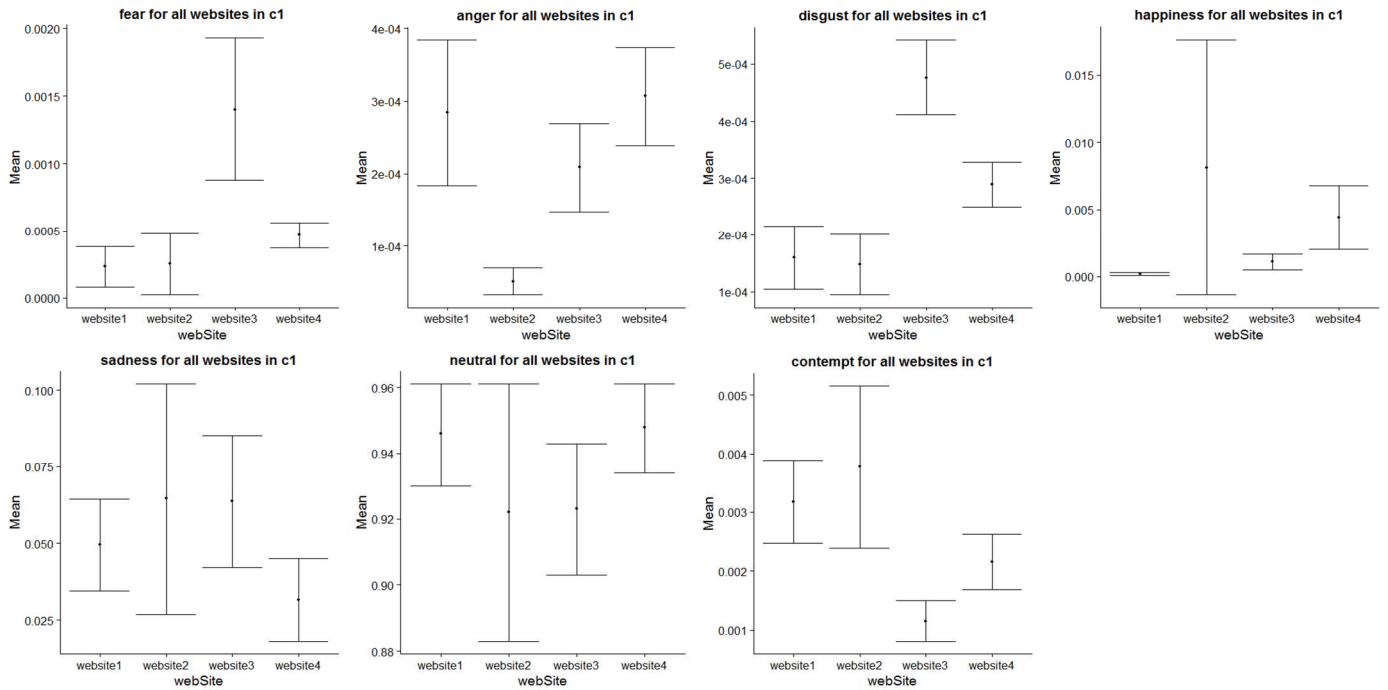


Figure 3- Emotion means and standard deviations for cluster 1 for women under 27 years

usability/aesthetics). We computed its mean and its confidence interval ⑤ for each emotion detected and each website version. We obtained 7 box-plot for each cluster (Fig. 3).

By analysing these clusters, 4 of them appeared to have not enough discriminating behaviours (see Table 2, Women > 27 in clusters 1 and 3 and Men > 27 in clusters 2 and 3, which are the strikethrough numbers). More than 83% of women's scores who are more than 27 years old are in the second cluster, making the two other ones not enough representative. In the group of men over 27 years, one website version is not represented in the data of cluster 2 (website 4) and cluster 3 (website 1). So, we did not use these 4 clusters from our results analysis in order to keep relevant results.

E. Results

1) Hypothesis 1

Our first research question is related to the possibility to categorize emotions depending on users' age and gender and on UI aesthetics and usability levels.

This first hypothesis is partly validated by our experimental results. Indeed, the identification of clusters by gender and age tends to show that it is possible to categorize emotions to obtain emotions behaviour depending on usability and aesthetics levels. However, 4 of our clusters are not relevant currently. More data would be required to show the relevance of all clusters. Anyway, the current relevant clusters are significant and can be used for hypothesis 2.

2) Hypothesis 2

The second hypothesis concerns the emotions thresholds for detecting usability and aesthetics problems from users' age and gender.

First of all, we need to identify emotions behaviours (in terms of one emotion or a combination of emotions) that can permit to identify a website version and so a usability and/or an aesthetics problem. For instance, we can see on Fig. 3 that the means for fear, disgust and contempt for website 3 (low usability and low aesthetics version, women under 27 years) are very different from the values obtained on the other websites for these same emotions. So, a low usability and a low aesthetics problem can be identified thanks to the detection of a certain level of fear, disgust or contempt for young women.

Secondly, we can see from Fig. 3 that each mean is given with a confidence interval. Here confidence intervals are of 95%, results can be reproduced in 95 percent of cases. This 5% of uncertainty can be used to determine the interval of values expected to determine a behaviour. For instance, the emotion behaviour studied previously is based on fear, disgust or contempt. The confidence interval for fear is between $8.79 \cdot 10^{-4}$ and $1.93 \cdot 10^{-3}$ (even if these numbers are close to 0, the difference is significant at 5% thresholds). These intervals provide us the emotions thresholds for identifying an UI problem. In our example, for women under 27 years, if fear is between $8.79 \cdot 10^{-4}$ and $1.93 \cdot 10^{-3}$, or disgust between $4.11 \cdot 10^{-4}$ and $5.42 \cdot 10^{-4}$ or contempt between $8.10 \cdot 10^{-4}$ and $1.5 \cdot 10^{-3}$, the corresponding website version can only be website 3. So, there are problems of aesthetics and usability.

To know if there is only a problem of usability (resp. aesthetics), we must study the behaviour of website 4, i.e., high aesthetics/low usability version, (resp. website 1, i.e., low aesthetics/high usability version). By analysing Fig. 3, we can provide conclusions for women under 27 years of age of cluster 1:

User category	Cluster	Website version	Emotions	Thresholds (min and max)	Problem
Women younger than 27	c1	website1	anger	$1.84*10^{-4}$ and $3.84*10^{-4}$	aesthetics
			& disgust	$1.04*10^{-4}$ and $2.16*10^{-4}$	
		website3	fear	$8.79*10^{-4}$ and $1.93*10^{-3}$	aesthetics and usability
			disgust	$4.11*10^{-4}$ and $5.42*10^{-4}$	
	contempt		$8.10*10^{-4}$ and $1.5*10^{-3}$		
	website4	disgust	$2.5*10^{-4}$ and $3.28*10^{-4}$	usability	
	c2	website1	anger	$4.08*10^{-4}$ and $1.02*10^{-3}$	aesthetics
			& fear	$4.78*10^{-4}$ and $1.11*10^{-3}$	
		website3	neutral	$9.11*10^{-1}$ and $9.26*10^{-1}$	aesthetics and usability
			contempt	$2.33*10^{-3}$ and $2.4*10^{-3}$	
		website4	anger	$2.49*10^{-3}$ and $4.47*10^{-3}$	usability
			happiness	$5.12*10^{-2}$ and $8.22*10^{-2}$	
	c3	website1	sadness	$3.05*10^{-1}$ and $3.61*10^{-1}$	aesthetics
			neutral	$6.02*10^{-1}$ and $6.64*10^{-1}$	
		website3	disgust	$1.72*10^{-3}$ and $2.22*10^{-3}$	aesthetics and usability
			sadness	$8.89*10^{-2}$ and $1.09*10^{-1}$	
			contempt	$1.63*10^{-3}$ and $1.99*10^{-3}$	
		website4	neutral	$8.26*10^{-1}$ and $8.67*10^{-1}$	usability
fear			$4.07*10^{-2}$ and $5.20*10^{-2}$		
anger			$2.96*10^{-4}$ and $3.76*10^{-4}$		
sadness			$1.14*10^{-1}$ and $1.5*10^{-1}$		
contempt		$5.09*10^{-3}$ and $8.6*10^{-3}$			
Women older than 27	c2	website1	happiness	$1.82*10^{-2}$ and $5.12*10^{-2}$	aesthetics
			contempt	$8.14*10^{-3}$ and $1.42*10^{-2}$	
	website3	anger	$3.25*10^{-3}$ and $4.82*10^{-3}$	aesthetics and usability	
	website4	fear	$2.46*10^{-4}$ and $5.78*10^{-4}$	usability	
		& happiness	$1.70*10^{-4}$ and $7.04*10^{-4}$		

Table 3 – Thresholds and UI problems detected for women

- Website 1 (low aesthetics version) is characterized by the level of anger (between $2.84*10^{-4}$ and $3.84*10^{-4}$) AND the level of disgust (between $1.04*10^{-4}$ and $2.16*10^{-4}$). A problem with aesthetics can thus be

identified if both emotions are detected between these levels.

- Website 4 (low usability version) can be identified by the level of disgust (between $2.5*10^{-4}$ and $3.28*10^{-4}$). So, a usability problem can be detected for young women, if emotions are inside these thresholds.
- Even if this does not help in identifying UI problem, we can notice that the version with high aesthetics and high usability (website 2) can be identified by the low level of anger (between $3.23*10^{-5}$ and $6.88*10^{-5}$).

We studied the 8 clusters in a similar manner. Table 3 for women and Table 4 for men summarize the results of this analysis. If it is the combination of several emotions values that discriminates, all emotions listed are required. This information is given by an “&” between emotions in the tables. Otherwise any of the listed emotions can be used to determine a problem.

We can notice that for men, there are some cases where no emotion data can discriminate an emotional behaviour. For men under 27 years old, for only one cluster, there is one behaviour that we cannot identify (cluster c2 and website3). For men older than 27 years, only one emotional behaviour is significant: this is the one related to problem in aesthetics and usability.

F. Discussion

The experiment allows us to partly validate our hypotheses: it shows that H1) it seems possible to categorize emotions depending on users’ age and gender and on UI aesthetics and usability levels and H2) some emotions thresholds can be used for detecting usability and aesthetics problems from users’ age and gender. As we choose confidence intervals of 95%, results can be reproduced in 95 percent of cases. Of course, an uncertainty of 5 percent remains. This is thus a first answer the lack identified in the related work: identifying potential issues to keep or improve the UI (H2) requires to establish links between UI quality factor (aesthetics and usability), emotions and user features (age and gender).

Although these results are encouraging, they must be mitigated as we designed this experiment considering very obvious differences between perceived aesthetics and perceived usability for distinguishing clearly these two factors. For instance, we carefully designed the low aesthetics version in order to avoid impacting also readability (guidance criteria) with inappropriate colours for labels. We also underline that: the tasks are non-critical and are without time and security constraints and that the experiment is limited to a single hand-crafted web-site. Finally, we notice that users can be inconsistent in their rational and emotional thinking [33], [34], can feel misleading emotions [35] and can vary in emotion arousal when facing the same stimuli [36]. It advocates for additional experiments to strengthen these results.

Lastly, the main limit of this work is that it does not allow us to fully cluster emotional behavior. Some clusters are not relevant and others do not discriminate enough emotions. Even if this is

User category	Cluster	Website version	emotions	Thresholds (min and max)	Problem
Men younger than 27	c1	website1	anger	$3.97*10^{-2}$ and $6.46*10^{-2}$	aesthetics
			disgust	$1.46*10^{-3}$ and $2.03*10^{-3}$	
			happiness	$6.32*10^{-4}$ and $2.41*10^{-3}$	
		website3	happiness	$2.68*10^{-4}$ and $4.97*10^{-3}$	aesthetics and usability
			& neutral	$9.05*10^{-1}$ and $9.24*10^{-1}$	
		website4	anger	$4.45*10^{-4}$ and $6.17*10^{-3}$	usability
	sadness		$1.20*10^{-1}$ and $1.55*10^{-1}$		
	c2	website1	anger	$1.01*10^{-3}$ and $3.57*10^{-3}$	aesthetics
		website3		no discriminant values	
		website4	contempt	$1.31*10^{-4}$ and $4.18*10^{-3}$	usability
	c3	website1	fear	$1.46*10^{-1}$ and $2.66*10^{-1}$	aesthetics
			anger	$1.83*10^{-2}$ and $3.27*10^{-2}$	
			disgust	$6.45*10^{-3}$ and $1.09*10^{-2}$	
			neutral	$5.82*10^{-1}$ and $6.38*10^{-1}$	
			contempt	$3.30*10^{-3}$ and $6.54*10^{-3}$	
		website3	fear	$1.7*10^{-2}$ and $3.09*10^{-2}$	aesthetics and usability
			happiness	$1.01*10^{-2}$ and $4.07*10^{-2}$	
			neutral	$6.98*10^{-1}$ and $7.48*10^{-1}$	
website4		sadness	$4.06*10^{-1}$ and $4.06*10^{-1}$	usability	
Men older than 27		c2	website1		no discriminant values
	website3		anger	$1.85*10^{-1}$ and $3.55*10^{-1}$	aesthetics and usability
	website4			no discriminant values	

Table 4 – Thresholds and UI problems detected for men

a limit, it does not challenge our approach. It simply implies that more data would be necessary to fully validate our hypotheses. However, results are currently good enough to be used in Perso2U, the prototype presented in the following section.

IV. INTREGRATION INTO PERSO2U

Perso2U is a prototype for adapting dynamically UIs to user's emotions.

A. Perso2U architecture

UI adaptation is based on the choice of the most appropriate variant for the current context of use (user, environment, platform). A *variant* is a variation of an UI created for a specific

context of use. For instance, there can be two different structures for a laptop and a smartphone, leading to two variants of the same UI.

Once a variant is chosen at runtime, some parameters can be applied to it. *UI Parameters* define a set of variables which personalize UI components (e.g., font-size, widgets, audio use or level, display and dialogues) regarding with identified context values (e.g., user's emotions). For instance, if the user's emotions are in the thresholds that correspond to an aesthetics problem then the UI parameter named background-color could be set to light-yellow. Then, these variables are sent to the displayed UI for applying the corresponding changes. Fig. 4 exemplifies this by showing different UI parameters applied to the same UI: here parameters are color and font size.

Based on the principles of variants and parameters, we proposed an architecture (Fig. 5) to adapt, at run-time, UI to the context of use including users' emotions [14]. This architecture is based on three components: the Inferring Engine ①, the Adaptation Engine ⑥ and the Interactive System ⑨. An automatic adaptation process starts from either (a) a need for a new UI to be displayed or (b) a change in the context of use. (a) can be exemplified by the user entering a web site: the home page has to be displayed. An example of (b) is the ambient light: when it increases, the contrast on the UI might be increased as well. In (b), the overall process is the following: the Inferring Engine ① monitors sensors ② to detect changes in the context of use. From these values, it deduces the new context of use dynamically. It embeds an Emotion Wrapper ③ which provides emotions values, coming from emotion detection tools ④ and included in the user model. The Inferring Engine ① sends ⑤ the computed context of use to the Adaptation Engine ⑥, which elicits accordingly a suitable UI variant and the UI parameters ⑩ for personalising it. Finally, the Interactive System ⑨ displays the variant and executes the changes related to the UI parameters. The whole process runs cyclically by following a time period defined by the designer.

B. Extension of the Perso2U inferring engine

The experiment described in section 3 shows that it can be possible to detect a negative emotion behaviour depending on user's gender and age. The Inferring Engine is in charge on computing this behaviour. The calculation relies on the user's emotions – detected at runtime by emotion detection tool and used as dynamic variables – and his/her age and gender – which can be declared by the user or deduced from pictures by the emotion detection tool and are used as static variables. Then the Inferring Engine sends the context of use with the user's emotions to the Adaptation Engine for adapting UI appropriately.

But to propose a relevant personalisation to emotions, the Adaptation Engine must also know if the problem is related to UI aesthetics or usability. So, the Inferring Engine must also send the identified problem source to the Adaptation Engine. This means that our prototype must be modified to add a parameter, the problem source that we name UI-problem, from the Inferring Engine to the Adaptation Engine.

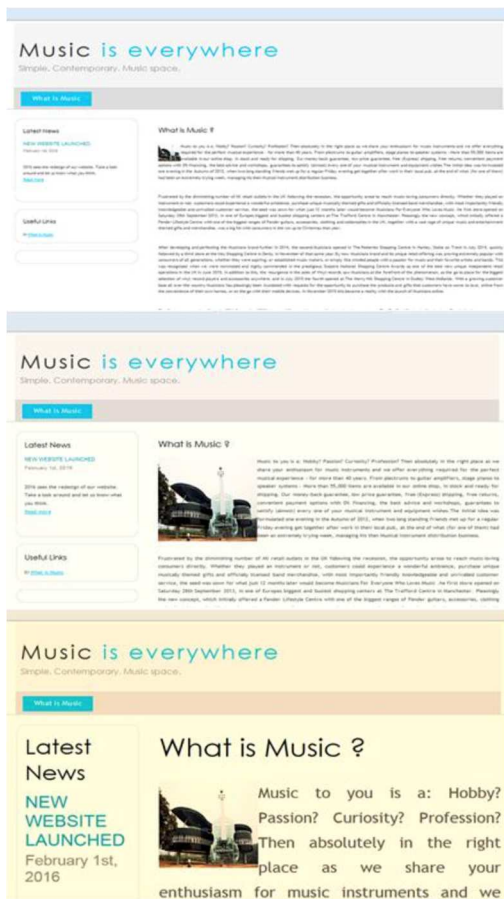


Figure 4. Parameter changes

From the experimental results, we can propose some inference rules about UI quality, that provide emotions (+, -, neutral) and the eventual problem.

Let's take as example: the case of women under 27 years of age for which the experimental results show that if fear is between $8.79 \cdot 10^{-4}$ and $1.93 \cdot 10^{-3}$, or disgust between $4.11 \cdot 10^{-4}$ and $5.42 \cdot 10^{-4}$ or contempt between $8.10 \cdot 10^{-4}$ and $1.5 \cdot 10^{-3}$, there are problems of aesthetics and usability. As any of the emotions detected in these intervals can discriminate the aesthetics and usability problem, the conditions about emotions thresholds are thus linked by an "or". When discrimination requires the combination of several emotions, conditions are linked by an "and". This is the case for the level of anger (between $2.84 \cdot 10^{-4}$ and $3.84 \cdot 10^{-4}$) and the level of disgust (between $1.04 \cdot 10^{-4}$ and $2.16 \cdot 10^{-4}$) that are both needed for discriminating a problem of aesthetics for women younger than 27 years old.

The implementation of this case in Perso2u is realized in two steps. First, Fig. 6 shows the thresholds definitions inside the inferring engine for women under 27 years. This section serves to include all needed thresholds variables previously defined in Tab. 2 and 3 without making distinction between clusters.

Secondly, a new function *checkUIQuality* was coded to identify UI quality problem from users' current detected emotions as illustrated in Fig. 7. Here, the Inferring Engine reads the static (gender, age) and dynamic (emotions) variables to make a comparison between the predefined thresholds and the current emotions. The conditions follow the criteria defined in Table 2 cluster 1.

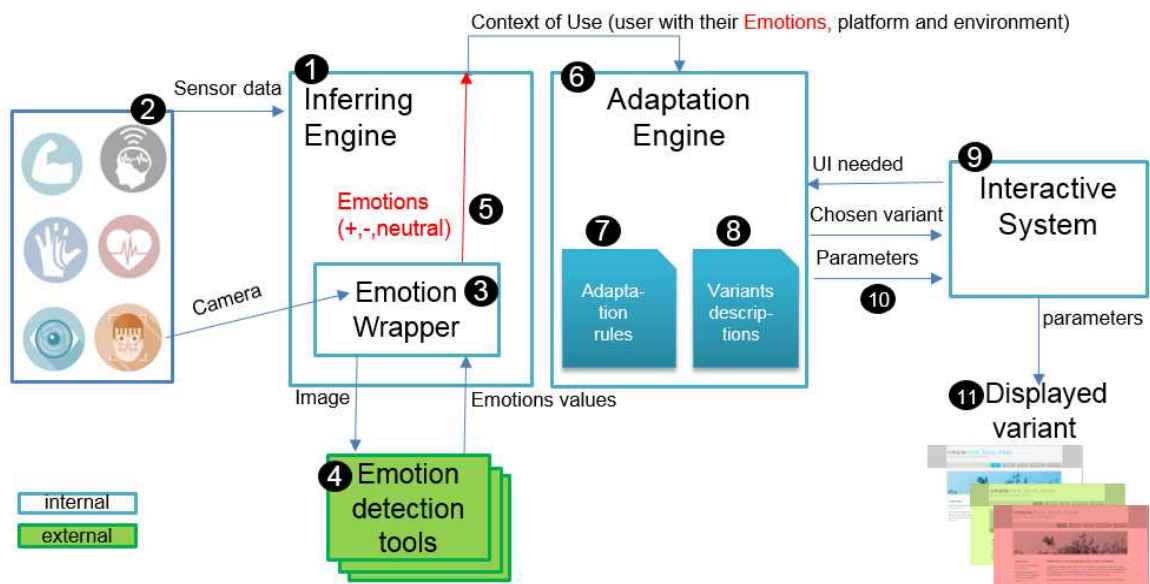


Figure 5. Global schema of the Perso2U architecture.

```

1 //Emotion Treshold for women < 27
2 //c1
3 //w1
4 //anger
5 var globalWc1w1Amin = 1.84e-4;
6 var globalWc1w1Amax = 3.84e-4;
7 //disgust
8 var globalWc1w1Dmin = 1.04e-4;
9 var globalWc1w1Dmax = 2.16e-4;
10 //w3
11 //fear
12 var globalWc1w3Fmin = 18.79e-4;
13 var globalWc1w3Fmax = 1.93e-3;
14 //disgust
15 var globalWc1w3Dmin = 4.11e-4;
16 var globalWc1w3Dmax = 5.42e-4;
17 //contempt
18 var globalWc1w3Cmin = 8.1e-4;
19 var globalWc1w3Cmax = 1.5e-3;
20 //w4
21 //disgust
22 var globalWc1w4Dmin = 2.5e-4;
23 var globalWc1w4Dmax = 3.28e-4;

```

Figure 6. Emotion thresholds definitions in the Inferring Engine for women lower than 27 years old.

Of course, other conditions are required for the other behaviours of Tables 2 and 3. Every condition results from the deduction of two variables: the user emotion (negative) and the UI quality problem (aesthetics, usability or aesthetics & usability). Finally, user emotion and UI-problem are stored in the context of use (user model).

Currently, detected emotions are used to identify some UI problems and for sending negative emotions to the Adaptation Engine. All emotions data that are not in the identified thresholds are considered as positive. A more precise algorithm could use the discriminant values for the website with high aesthetics and high usability to detect positive emotions and keep the neutral values for all data that do not permit discrimination. For instance, in section 3, we have identified that the website with high aesthetics and high usability can be discriminated by a low level of anger for a cluster of women under 27 years of age. So, the low level of anger could be interpreted as a positive emotion. Currently positive emotions are not used by the adaptation engine and no distinction is made between positive and neutral emotions.

```

1 function checkUIQuality() {
2   if (contextUse.user.gender == 'w') {
3     if (contextUse.user.age < 27) {
4       //c1w1
5       if (globalAngerValue > globalWc1w1Amin &&
6           globalAngerValue < globalWc1w1Amax &&
7           globalDisgustValue > globalWc1w1Dmin &&
8           globalDisgustValue < globalWc1w1Dmax) {
9         contextUse.user.emotion = '-'; //negative
10        globalUIproblem = 'A'; //aesthetics
11      } else {
12        //c1w3
13        if ((globalFearValue > globalWc1w3Fmin &&
14            globalFearValue < globalWc1w3Fmax) ||
15            (globalDisgustValue > globalWc1w3Dmin &&
16             globalDisgustValue < globalWc1w3Dmax) ||
17            (globalContemptValue > globalWc1w3Cmin &&
18             globalContemptValue < globalWc1w3Cmax)) {
19          contextUse.user.emotion = '-'; //negative
20          globalUIproblem = 'AU'; //aesthetics & usability
21        } else {
22          //c1w4
23          if (globalDisgustValue > globalWc1w4Dmin &&
24              globalDisgustValue < globalWc1w4Dmax) {
25            contextUse.user.emotion = '-'; //negative
26            globalUIproblem = 'U'; //usability
27          }
28        }
29      }
30    }
31  }
32 }
33 }

```

Figure 7. Inferring Engine Rule to find a UI quality issue.

The execution of the proposed implementation in the global architecture (Fig. 5) produces the following process. First, user's gender and age are stored in the context of use in the user model. Then, while the user interacts with the UI, the emotion wrapper ⑤ retrieves the detected emotions from the emotion recognition tool (Microsoft Emotion API). From detected emotions, the Inferring Engine ① calls the function *checkUIQuality* to look for predefined negative behaviours. Then, if a negative behaviour is detected (i.e. if emotions values followed the conditions exposed in Table 3), user emotions and UI-problem parameters are modified and sent with other contextual elements to the Adaptation Engine. The Adaptation Engine ⑥ reads the UI-problem and the user-emotions and can trigger an appropriate UI adaptation.

# Image	User				UI			Detected Emotions							Expected		Observed			Test results
	UserId	Age	Gender	image	Web	minute	second	neutral	fear	anger	sadness	disgust	contempt	happiness	Emotion	UI issue	Emotion	UI issue	cluster	
1	1	26	W	18/95/1.jpg	w1	45	13	0,612451136	0,004084732	0,0003121	0,3783435	0,0001941	0,0020064	0,0006083	negative	A	negative	A	c3	OK
2	1	26	W	18/95/2.jpg	w1	45	22	0,7462314	0,000322715	0,0001651	0,0268834	0,0001811	0,1816033	0,0432199	positive	n.a.	positive	n.a.	n.a.	OK
3	1	26	W	18/95/3.jpg	w1	45	32	0,7428729	0,000623453	0,0005082	0,0414263	0,0005109	0,1683542	0,0433454	negative	A	negative	A	c2	OK
4	1	26	W	18/95/4.jpg	w1	45	42	0,721467733	0,000603723	0,0005851	0,0605682	0,0007054	0,192517	0,0218351	negative	A	negative	A	c2	OK
5	1	26	W	18/95/5.jpg	w1	45	52	0,729369342	0,000708876	0,0007718	0,0372388	0,0008136	0,1889769	0,039334	negative	A	negative	A	c2	OK
6	1	26	W	18/95/6.jpg	w1	46	2	0,8091888	0,001393284	0,0007798	0,0555846	0,0005829	0,110695	0,0194708	positive	n.a.	positive	n.a.	n.a.	OK
7	1	26	W	18/95/7.jpg	w1	46	12	0,536161959	0,003813865	0,0013723	0,4536556	0,0005061	0,0026888	0,0002076	positive	n.a.	positive	n.a.	n.a.	OK
8	1	26	W	18/95/8.jpg	w1	46	22	0,4728568	0,005632889	0,0010341	0,5144534	0,0005723	0,0039513	0,0003561	positive	n.a.	positive	n.a.	n.a.	OK
9	1	26	W	18/95/9.jpg	w1	46	32	0,6986122	0,003568676	0,0014565	0,293213	0,0004421	0,00125	0,0003219	positive	n.a.	positive	n.a.	n.a.	OK
10	1	26	W	8/95/10.jpg	w1	46	42	0,75827986	0,007298846	0,0006054	0,2297023	0,0002139	0,0019421	0,0003944	positive	n.a.	positive	n.a.	n.a.	OK

Table 5 – Ten first results of inferring rules tests for a woman less than 27 years old.

C. Testing of the new inference rules in Perso2u

The implementation of the rules for inferring negative emotions contains a lot of conditions which correspond to the verification of the thresholds defined by the statistical analysis. However, the implementation removes the clusters of the analysis, as the Inferring Engine cannot know in advance which cluster a person belongs to. So, we need to test carefully if the results of the Inferring Engine are correct.

The test aims at verifying if a correct emotion and UI issue is provided to the Adaptation Engine by using the inference rules. The inference rule for women less than 27 years old was tested with 105 images of three users. The test was done by comparing the expected and observed values (inferred and sent to the Adaptation Engine). These values include the inferred emotion (negative, positive) and the UI issue: aesthetics (A), usability (U) or aesthetics & usability (AU) or not available (n.a.) when a negative emotion is not found. To illustrate the results of the test of the inferring rule algorithm, Table 5 shows ten results for one user. The “Test results” column shows if the observed value matches the expected one (OK) or not (not OK). For instance, for image #1 in Table 4, has expected values {negative, A} (emotion is negative, there is an aesthetics problem), while its observed values are {negative, A}, that correspond exactly to the expectations.

The tests allowed us to identify 4 errors corresponding to thresholds badly reported in tables 3 and 4. The code of the inferring rules was corrected accordingly. Now the code is robust and the Inferring Engine is able to deduce a negative emotion and UI issue according to Tables 3 and 4.

V. CONCLUSIONS AND PERSPECTIVES

This paper presents the results of an experiment and states that 1) it can be possible to categorize emotions depending on users' age and gender and on UI aesthetics and usability levels and that 2) some emotions thresholds can be defined for detecting usability and aesthetics problems from users' age and gender. These results are used in a UI adaptation system to infer negative emotions and UI problems based on users' emotions, age and gender. Adaptations triggered by emotions are expected to improve UX.

Other experiments could strengthen the current results. An experiment, similar to the one described in this paper with more participants, could be realized to characterize other users' clusters. This would permit to obtain a more precise adaptation system. Another experiment could validate the proposed inference rules.

It would also be interesting to strengthen the statistical analysis to obtain a statistical model, which would allow us to obtain a more precise determination of UI problems. Indeed, a logistic regression can estimate the probability being in front of a type of website (e.g. low usability or low aesthetics) depending on the gender, the age and the detected emotions. This model is interesting because it makes it possible to obtain the level of confidence of the prediction, while in our rules the thresholds

are at 5% level. The statistical model is more precise than the rules model.

Currently, our prototype proposes simple adaptations (colour and text font) to emotions without referring to relevant work in psychology concerning preferences by gender or age. Coupling this existing knowledge with our system would allow us to provide a more relevant adaptation system.

Moreover, our experimental results can also be used in UI evaluation as they allow a system to detect some usability or aesthetics problems. This opens a large spectrum of possibilities for UI evaluation based on emotions. However even if lots of questions remain opened, this work is a first step toward providing better UX thanks to emotions.

ACKNOWLEDGMENT

This work has been supported by the Labex PERSYVAL-Lab in the University Grenoble-Alpes.

REFERENCES

- [1] M. Winckler, C. Bach, and R. Bernhaupt, “Identifying user experience dimensions for mobile incident reporting in urban contexts,” *IEEE Trans. Prof. Commun.*, vol. 56, no. 2, pp. 97–119, 2013.
- [2] R. Bernhaupt, “User experience evaluation in entertainment,” in *Evaluating user experience in games*, Springer, 2010, pp. 3–7.
- [3] N. Tractinsky, A. S. Katz, and D. Ikar, “What is beautiful is usable,” *Interact. Comput.*, vol. 13, no. 2, pp. 127–145, 2000.
- [4] A. N. Tuch, S. P. Roth, K. Hornbæk, K. Opwis, and J. A. Bargas-Avila, “Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI,” *Comput. Hum. Behav.*, vol. 28, no. 5, pp. 1596–1607, Sep. 2012.
- [5] A. De Angeli, A. Sutcliffe, and J. Hartmann, “Interaction, usability and aesthetics: what influences users' preferences?,” in *Proceedings of the 6th conference on Designing Interactive systems*, 2006, pp. 271–280.
- [6] D. Stone, C. Jarrett, M. Woodroffe, and S. Minocha, *User interface design and evaluation*. Morgan Kaufmann, 2005.
- [7] W. O. Galitz, *The essential guide to user interface design: an introduction to GUI design principles and techniques*. John Wiley & Sons, 2007.
- [8] G. Calvary et al., “Plasticity of user interfaces: A revisited reference framework,” in *In Task Models and Diagrams for User Interface Design*, 2002.
- [9] F. Paterno, “User Interface design adaptation,” *Encycl. Hum.-Comput. Interact.* 2nd Ed, 2013.
- [10] D. Cyr and C. Bonanni, “Gender and website design in e-business,” *Int. J. Electron. Bus.*, vol. 3, no. 6, pp. 565–582, 2005.
- [11] J.-B. E. Steenkamp and I. Geyskens, “How country characteristics affect the perceived value of web sites,” *J. Mark.*, vol. 70, no. 3, pp. 136–150, 2006.
- [12] M. Bakaev, K.-H. Lee, and H. I. Cheng, “The aesthetic and emotional preferences of the elderly and the design factors for e-business web sites,” in *The Eighth Pan-Pacific Conference on Occupational Ergonomics (PPCOE 2007)*, 2007.
- [13] S. Djasasbi, M. Siegel, and T. Tullis, “Generation Y, web design, and eye tracking,” *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 5, pp. 307–323, 2010.
- [14] J. Galindo, S. Dupuy-Chessa, and É. Céret, “Toward a UI Adaptation Approach Driven by User Emotions,” presented at the ACHI07, Nice, France, 2017.
- [15] K. Hook, “Affective Computing,” *The Interaction Design Foundation*. [Online]. Available: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/affective-computing>. [Accessed: 06-Apr-2018].

- [16] R. W. Picard, "Affective computing," 1995.
- [17] P. Ekman, "An argument for basic emotions," *Cogn. Emot.*, vol. 6, no. 3–4, pp. 169–200, 1992.
- [18] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant," *J. Pers. Soc. Psychol.*, vol. 76, no. 5, p. 805, 1999.
- [19] K. Forbes-Riley and D. Litman, "Designing and evaluating a wizarded uncertainty-adaptive spoken dialogue tutoring system," *Comput. Speech Lang.*, vol. 25, no. 1, pp. 105–126, Jan. 2011.
- [20] K. Porayska-Pomsta, M. Mavrikis, and H. Pain, "Diagnosing and acting on student affect: the tutor's perspective," *User Model. User-Adapt. Interact.*, vol. 18, no. 1–2, pp. 125–173, 2008.
- [21] A. C. Graesser et al., "The relationship between affective states and dialog patterns during interactions with AutoTutor," *J. Interact. Learn. Res.*, vol. 19, no. 2, p. 293, 2008.
- [22] A. Esposito, "Affective Signal Processing (ASP): Unraveling the mystery of emotions, by Egon L. van den Broek," *J. Ambient Intell. Smart Environ.*, vol. 4, no. 1, pp. 67–69, 2012.
- [23] J. H. Janssen, E. L. Van Den Broek, and J. H. Westerink, "Tune in to your emotions: a robust personalized affective music player," *User Model. User-Adapt. Interact.*, vol. 22, no. 3, pp. 255–279, 2012.
- [24] S. Meudt et al., "Going Further in Affective Computing: How Emotion Recognition Can Improve Adaptive User Interaction," in *Toward Robotic Socially Believable Behaving Systems - Volume I*, vol. 105, A. Esposito and L. C. Jain, Eds. Cham: Springer International Publishing, 2016, pp. 73–103.
- [25] E. Hudlicka and M. D. Mcneese, "Assessment of user affective and belief states for interface adaptation: Application to an Air Force pilot task," *User Model. User-Adapt. Interact.*, vol. 12, no. 1, pp. 1–47, 2002.
- [26] F. Nasoz, "Adaptive intelligent user interfaces with emotion recognition," University of Central Florida Orlando, Florida, 2004.
- [27] C. Märtin, S. Rashid, and C. Herdin, "Designing Responsive Interactive Applications by Emotion-Tracking and Pattern-Based Dynamic User Interface Adaptation," in *International Conference on Human-Computer Interaction*, 2016, pp. 28–36.
- [28] S. Dupuy-Chessa, Y. Laurillau, and E. Céret, "Considering aesthetics and usability temporalities in a model based development process," in *Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine*, 2016, pp. 25–35.
- [29] S. Mann, *Emotion: All That Matters*. Hachette UK, 2014.
- [30] S. J. Simon, "The impact of culture and gender on web sites: an empirical study," *ACM Sigmis Database*, vol. 32, no. 1, pp. 18–37, 2000.
- [31] E. Ozdemir and S. Kilic, "Young Consumers' Perspectives of Website Visualization: A Gender Perspective - ProQuest," 2011. [Online]. Available: <https://search.proquest.com/openview/23ce2c57134808a5356dc26c843aa3e8/1?pq-origsite=gscholar&cbl=216138>. [Accessed: 21-Mar-2018].
- [32] G. Moss, R. Gunn, and J. Heller, "Some men like it black, some women like it pink: consumer implications of differences in male and female website design," *J. Consum. Behav.*, vol. 5, no. 4, pp. 328–341, 2006.
- [33] *Cognitive dissonance: Progress on a pivotal theory in social psychology*, vol. xviii. Washington, DC, US: American Psychological Association, 1999.
- [34] L. Festinger, *A Theory of Cognitive Dissonance*. Stanford University Press, 1962.
- [35] P. Goldie, "Explaining expressions of emotion," *Mind*, vol. 109, no. 433, pp. 25–38, Jan. 2000.
- [36] D. Hume, "Emotions and moods," in *Organizational behavior*, 2012, pp. 258–297.