The Use of Neurometric and Biometric Research Methods in Understanding the User Experience During Product Search of First-Time Buyers in E-Commerce

Tuna Çakar[1(✉)], Kerem Rızvanoğlu[2], Özgürol Öztürk[2], Deniz Zengin Çelik[2], and İrfan Gürvardar[3]

[1] Big Data Analytics Program, MEF University, İstanbul, Turkey
tuna.cakar@tunasc.com
[2] Faculty of Communication, Galatasaray University, İstanbul, Turkey
krizvanoglu@gmail.com, ozgurol@gmail.com, denizzengincelik@gmail.com
[3] Gittigidiyor (Ebay Turkey) Inc. Co., İstanbul, Turkey
igurvardar@ebay.com

Abstract. Understanding user experience (UX) during e-commerce has been a relatively important research area especially in the last decade. The use of conventional methods in UX such as task-observation, in-depth interviews and questionnaires has already contributed for the measurement of the efficiency and effectiveness. This empirical study has aimed to make use of both conventional and neuroscientific methods simultaneously to provide a richer analysis framework for understanding the product search experience of the first-time buyers. The current work provides insights for the results from the combined use of conventional and neuroscientific-biometric methods in a UX study. Although this has been an exploratory study within a limited literature, the obtained results indicate a potential use of these methods for UX research, which may contribute to improve the relevant experience in various digital platforms.

Keywords: User experience (UX) · E-commerce · Product search · Decision making · Traditional user research methods · Neuroscientific methods · Neuroergonomics

1 Introduction

User experience (UX) is defined as "a person's perceptions and responses that result from the use or anticipated use of a product, system, or service" (ISO 9241-210 (2009)). This definition is directly related to the inner states of the experiencer in terms of both cognitive and affective dimensions. On the other hand, it covers the perception, evaluation and decision related to the product in use. It has been important to understand this integrated experience by examining different research methods and comparing these findings. The conventional methods including the surveys/questionnaires, think-aloud procedures and in-depth interviews have contributed considerably for

A. Marcus and W. Wang (Eds.): DUXU 2017, Part I, LNCS 10288, pp. 342–362, 2017.
DOI: 10.1007/978-3-319-58634-2_26

understanding the problems during the use of internet sites for shopping. On the other hand, the use of neuroscientific and psychophysics methods, also mentioned as neuroscientific and biometrics in this text, has also grasped attention with the exciting idea of providing objective means of understanding cognitive and affective processes during the user experience under online environments (Chai *et al*. 2014). Despite the fact this idea has also strong limitations such that an attached device is usually necessary, it has been of interest for many researchers to explore such potential use in this interdisciplinary area of research. This empirical and exploratory study has aimed to understand the user experience of first-time buyers during e-shopping and specifically during product search in e-commerce via the use of neurophysiological and psychophysics research methods as well as the conventional methods. The main research scope is about understanding the potential negative factors, specifically obstacles, during e-shopping processes by both neurophysiological and conventional methods. We have also aimed to compare and to contrast the consistencies and differences between the findings of conventional and neurophysiological-biometric methods.

2 Theoretical Background

The current academic literature has provided two subcategories regarding the research on user experience evaluation methods. These are the subjective and objective evaluation methods. The former one generally relies on the self-reported data about the digital platforms presented whereas the latter one is not targeted at the perception of the user but rather the performance metrics such as total time spent on a web-site or the number of mouse clicks at a specific online medium allocate the main scale that is quantifiable and/or measurable (Hornbæk 2006). The objective measures have generally been referred to the use of eye-tracking method with the outputs such as number of fixations, total time spent on a specific Area of Interest (AOI). All these evaluation methods have been regarded as conventional user experience research methods by which generally a combination of these is preferred during research. Despite the fact these have been the dominant research methodologies used in UX research and professional projects, there are alternative techniques recently emerged involving neurophysiological and biometric/psychophysics methods. The use of neuroscientific and biometric methods has generally been referred to as neuromarketing or neuroergonomics but the former term covers different fields of application than user experience such as the assessment of TV advertisements, visual design tests, brand-adjective and in-store tests. The initial expectation from neuroscientific methods has been to provide a revolutionary change from conventional methodologies to neurophysiological and biometric methods. However, the obtained research findings obtained via conventional and neurophysiological methods has revealed that these methods should better be used as complementary methods to provide a more fruitful framework.

2.1 First-Time User Experience (FTUE)

The initial stages/steps of user experience on a software or an internet site has been defined as first-time user experience that involves signing up, the other configuration steps. Although one user might have an extensive experience with a similar product, he/she will have a FTUE with a new product regardless of how his/her previous experience will have an impact on this FTUE (Drenner et al. 2008). For assessing the usability of an e-commerce site, it is important to observe the behavior of first-time users/buyers who do not have any knowledge about the content/format of this web-site. Time effectiveness is one of the key factors for establishing a successful and long-term relationship between user and the service during FTUE. Time effectiveness has been defined as the effective use of time by the user to complete the given task within a considerably accepted range (Drenner et al. 2008). Thus, it inevitably depends on the complexity of the given task as well as the individual differences. The second factor is intuitiveness that refers to the usability of the presented software without any guidance that means the user can intuitively figure out what to do next without a need for external help. Thus, it is possible for researchers to assess several different factors including layouts, graphics, location of buttons and other features over the experiments done with FTUE to develop more user-friendly designs and to allocate the potentials problems/obstacles. The third factor is patience.

2.2 Subjective Measures

There are many different user evaluation methods including surveys, interviews, and contextual inquiry frequently used in user experience research done both in professional and academic studies (Law *et al*. 2009). Each method has different strengths and weaknesses but they are generally focused on obtaining information directly from the participants in which the verbal declarations are the main sources of information. Understanding the affects experienced by the users have been issued by Boehner as a critical component of the subjective interpretation and highly important for HCI research domain (Boehner et al. 2005). The emotional responses of the users have been collected retrospectively and directly from the users by which they are expected to state how they have interacted with the presented material/media. Although this is a common use in UX studies, it is widely accepted that the self-reports and other subjective measures are generally directed towards the best and worst experiences rather than providing a full account of the experience. Moreover, it has also been a critical issue that the participants could be influenced by the experimenters' demand so that they might not be delivering the correct information. For instance, the participants can positively be motivated towards the presented web-site with the belief that the experimenters are the designers of it thus they might not be glad to hear/receive negative feedback during the usability and user experience test. In addition to this, it has also been issued that participants might be willing to give the politically correct replies rather than giving the correct information that is not in line with the social norms (Ravaja 2004).

In-depth Interviews. In-depth interviews are mainly conducted for exploring the ideas, comments, perspectives of a small groups of people/participants on a given situation. In this respect, it was important to obtain the experiences, expectations, thoughts and suggestions of the participants about the operations, processes, and outcomes related to this e-commerce site. This type of detailed information composed of suggestions and expectations are useful in the sense to provide further explorations in more detail. The insights from in-depth interviews are commonly used to provide relevant context for understanding the data from other domains or sources. In-depth interviews are preferred over surveys, since they provide much more detailed information. Moreover, it might be preferable for participants when compared to surveys since they are more likely to be a relaxed and conversation-like way of data collection. However, there are also methodological limitations of this method that should be handled with care. First, the obtained responses from the interviews might be biased, since the participants might be looking for their own benefits. Secondly, the conduction of the interviews, transcription stage, and the evaluation of the responses might take too much time when compared to the analysis durations of the other methods. Thirdly, it is crucial to comfort the participant in the sense by an experienced interviewer that the participant could easily tell her opinions about the product or web-site of target. Moreover, the interviewer should not ask leading questions or should not give clues about his personal opinions and should fully attend the participant's interview. Fourthly, it is significant to recognize that the obtained results are mostly not generalizable due to the fact the sample size is generally small and random sampling methods are not performed. Thus, one should be precautious about reporting and presenting the obtained results and related conclusions.

Surveys. Surveys are tools for collecting information from the users about the experience they had while interacting with a website or software program. These tools contain set of questions to understand one's opinions, preferences, attitudes, and characteristics on a given item. The responses from the users will be directing the professionals or researcher to provide a more user-friendly version of the product or the service. The responses from a subset of the target population might be applied to a broader population. For the e-commerce research, there are three main issues regarding the usability of the website: (1) perceived ease-of-use, (2) satisfaction, and (3) perceived time.

Task Observation. Task observation is a frequently used research method that investigates the user during the experience of the presented product or service via online (or offline) observations by an expert. The expert identifies the potential factors that lead to the problems for instance during the use of the website in this context. These identifications regarding the UX could then be validated through the in-depth interviews and/or application of surveys. The task observation method is also critical and useful for understanding the implicit factors that could not be noticed by the user herself but the field expert can observe and identify the potential problems. These accounts from the observations are potentially beneficial for improving the usability of the target product or service.

There are a couple of neurophysiological and biometric methods that have increasingly been used by researchers as well as professionals in the field. The neurophysiological methods are the ones that collect data directly from the brain but the biometric methods have an indirect access to the brain activity. The most widely used neurophysiological methods include EEG/ERP, fMRI and fNIRS (optic brain imaging). On the other hand, the biometric methods involve galvanic skin response (GSR), eye-tracker (E-T), and heart rate (HR). In this study, EEG/ERP, eye-tracker and GSR methods were used for data collection. EEG/ERP method has generally been used for the valence/intensity of the experienced pleasantness (affect) with the calculation of Frontal Alpha Asymmetry (FAA). GSR has been used to detect the changes in the arousal dimension by measuring the electrical resistance of the skin. Eye-tracker has been used to keep track of the eye-movements of the participants thus the use of this method provided the potential explanations for the changes in neuroscientific and biometric indexes. The objective measures (as opposed to the subjective measures) are argued to have the advantage of providing continuous and quantitative output during the user experience. Moreover, the objective measures have also the advantage of not being dependent on any language (or any other verbal articulation). On the other hand, there are a couple of disadvantages of the objective methods such as the variations between subjects (due to individual differences), determination of the significant events especially in large datasets, and providing interpretations for the determined significant events (since different mental states might result with similar outputs). These pose challenging situations for the researchers making use of objective measures while trying to construct a reliable and replicable framework.

Time on Task (Reaction Time) Measure. Time on Task (ToT) measure has been a common measure specifically in UX research to infer/understand the easiness/difficulty of the given task (REF). Total time spent on a given task (for instance, the signing up process to a web-site) were calculated through the intervals between the initiation and completion of a given task. This measure is especially important to understand if a user spends excessive amount of time to complete a task.

The Valence of the Emotional State: Positiveness (FAA). The current academic literature indicates that frontal alpha asymmetry (FAA) can be used as a potential indicator for identification of the valence of emotions as depicted in several basic empirical studies (Briesemeister 2013). According to Davidson's model (1979), there is a hemispheric asymmetry regarding the alpha oscillations among the frontal channels. This neurophysiological model implicates that the left frontal cortex is more activated during the processing of positive affects whereas the left prefrontal cortex is responsible for the negative affect experienced (Davidson *et al*. 1979). This fundamental has extensively been used in neuromarketing and consumer neuroscience research. One of the main studies in the academic literature has targeted to address how the frontal alpha asymmetry can be used to dissociate between three similar versions of TV ads from the same brand (Ohme *et al*. 2009). The FAA approach has been used in wide range of applications from product design tests to TV advertisement assessments and to flavor tests (Tomico *et al*. 2008). One of the newest research and application domain has been suggested to understand the emotional states during user experience. The limited empirical evidence so far indicates that FAA can be a potential and useful index in

human-computer interaction studies (Chai *et al*. 2014). As stated in the previous section, the results of the empirical study done by Chai *et al.* (2014) has shown that excellent user experience is correlated with a positive trend in FAA. On the other hand, it has also been revealed that negative emotional states have been correlated with the negative trends in FAA (Wheeler *et al*. 1993). One of the main neurophysiological paradigms of the current study is based on the use of FAA index through the inspection of first-time buyers during e-commerce. It was initially hypothesized that the participants will have negative peaks as they have difficulties during a given e-commerce task.

Arousal dimension of the Emotional states: GSR. Arousal has been another crucial dimension for determining the emotional state of the participants. The arousal level has been shown to be correlated with the galvanic skin response measure (GSR) that measures the electrical resistance of the skin through giving very low levels of amperes (at milliampere level) and measuring the voltage (Boucsein 2012). GSR is a psychophysics method that has been mentioned with different names in the literature as such: electrodermal response (EDR), psychogalvanic reflex (PGR), and skin conductance resistance (SCR). The GSR method has already shown to be correlated with sympathetic activity that is also demonstrated to be correlated with the emotional arousal (Westerink *et al*. 2008). Since GSR has been associated with the level of emotional arousal regardless of the intensity/direction of the emotion experienced, it has also been argued that various emotional states including fear, anger, orienting response might cause similar GSR reactions (Carlson 2013). On the other hand, it might be possible to infer about the processing difficulties of the participants relying on the signal peaks acquired from the GSR device (Carlson 2013). Arousal dimension has been crucial for this kind of studies mainly because of the activator aspect of the arousal such as being as a pusher for the fight-or-flight actions. In this respect, the detected peaks in the arousal dimension could be inferred as the instantaneous processing problems during e-commerce. However, one needs to be precautious during such inferences, since these peaks might also be caused by positive affect. In this respect, the findings from the FAA should also be evaluated as well as the contributions from the subjective measures.

2.3 Research Questions and Purpose

The main purpose of this study is to present and discuss the findings obtained by several different methods, which focus on understanding the product search experience of the first-time buyers in a desktop e-commerce web site. This empirical study mainly revolves around two main research questions: (1) What are the similarities/overlaps of the findings/results from conventional and neuroscientific/biometric methods? (2) What are the inconsistencies/differences and problems regarding the obtained results from these methods?

3 Methodology

This empirical study focuses on understanding the online shopping experience in desktop marketplace e-commerce website through the neuroscientific/biometric methods including EEG/ERP, eye-tracker, galvanic skin response (GSR), pulse rate (PR) as well as UX research methods such as task-observation, in-depth interview and survey.

3.1 Participants

The participants were all males, 24–35 years of age, right-handed, and actively working as professionals at least for 3 years. The participants were also required not to have done any online shopping via a specific web-site before. This requirement was significant to assess their first-time user experience (FTUE) in this specific web-site. They were also required not to use any psychiatric drugs in the last 6 months. The participants were informed about the aim and content of the experiment before the session starts. They were reminded that if they feel uncomfortable during the experiment, they might leave the experiment without any hesitation. The ethics approval for this study has been acquired from the Ethics Committee of Acıbadem University (Istanbul, Turkey).

3.2 Procedure

The participants were asked to sign up to this web-site, they were then to choose a specific item category to buy, to provide three concrete options for this category and then to purchase one of these options. These participants were motivated by a gift card of 250 TL (approximately 73 USD) as a contribution to their shopping task. In the post-test session, the participants were given a survey about the experiment and invited to an in-depth interview through a retrospective think-aloud protocol to be able to grasp their online shopping experience in more detail. The participants were instructed about the experimental task in detail.

The experimental session is composed of 4 different tasks: (1) signing in, (2) product search, (3) find 3 items of their interest & add to cart, (4) purchase one of these 3 items. The first task was that the participants were to login to this e-commerce site without any time limitation. For the second task, the participants were instructed to find a product category of their interest and they were asked to find three alternatives for this product category. Moderators reminded that it was important to reach the ticket limit of 250 TL to be able to activate the given tickets.

3.3 Conventional UX Research Methods

Test Evaluation. *Task-Observation.* A team of two researchers were responsible for the observation of the participants during task execution: A moderator and a facilitator. The test was led by a moderator and the task-execution was directly observed and recorded on a structured observation sheet by both the moderator and a facilitator.

Post-test Evaluation. *In-depth Interviews.* After the online recordings, the participants participated an in-depth interview by which their introspective judgments about their experience were asked and their replies were noted down. The participants were shown the related fields on the screen to enable them remember and actively participate in the interview session. The questions were generally centered at getting more about the experience of the participants in detail.

Surveys. A user experience questionnaire has been performed just after the in-depth interview session. The participants were asked to rate the perceived quality of their experience regarding different aspects. In this survey, three main aspects were questioned: (1) perceived ease-of-use, (2) satisfaction, and (3) perceived time. The participants were orally asked to rate the given aspect in 5-point Likert scale. The responses were noted down by the moderators.

3.4 Neurophysiological and Biometric Research Methods

Time-on-Task (ToT) Measure. Time-on-Task (ToT) measure has been a common measure specifically in UX research to infer and understand the degree of difficulty of the given task. ToT measures were calculated through the intervals between the initiation of a given task and completion of that task.

EEG Recording and Data Analysis. A 32-channel wireless dry-electrode EEG/ERP system, has been used to calculate Frontal Alpha Asymmetry (FAA) that has been claimed to be an indicator of approach/withdrawal tendencies of the participants. Thus, the valence dimension of the emotional experience has been estimated by these empirical outputs. The main scope has been to capture the instants with negative peaks including the issues related to personal information entry of the email addresses, log in to the email accounts, identity number and the issues related to the usability of the website including delay in the page loadings and unselecting items in the shopping cart.

The EEG data was continuously recorded from 32 scalp sites using dry electrodes mounted in an elastic cap arranged according to the 10–10 international placement system (g.Nautilus). The online reference was right mastoid (A2). The impedances of the EEG electrodes were below 50 kΩ. The EEG data were amplified with a bandpass filter of 0.01–45 Hz and digitized at 500 Hz. The recordings obtained from the prefrontal and frontal regions of the cortex (Fp1, Fp2, AF3, AF4, F3, F4, F7, and F8) were analyzed (Fig. 1).

Fig. 1. Wireless g.Nautilus 32-channel EEG/ERP device

The EEG data was processed using MATLAB 2014b and Fieldtrip. All the EEG data were DC corrected and re-referenced to linked mastoids offline. The filter was set to 40 Hz low pass and 0.1 Hz high pass. Then, the EEG data were epoched into periods of 500 points (i.e., 1000 ms). The power of alpha band (8–12 Hz) for each of the recorded electrodes was calculated for further analysis.

According to the previous empirical studies in the relevant literature (Ohme *et al.* 2010), the frontal alpha asymmetry (FAA) index was calculated as the difference between right-hemispheric data minus left-hemispheric data with the following formula (Davidson *et al.* 1979):

$$\ln(\text{right alpha power}) - \ln(\text{left alpha power}) \quad (1)$$

Due to the negative correlation between alpha power and brain activation, the positive score of the FAA index implies the dominance of left PFC and the negative score of the FAA index implies the dominance of right PFC. These neurophysiological and psychophysics outputs have been plotted in a video synchronized with screen recorder and eye-tracker. The obtained data has been statistically analyzed to find the significant peaks within the whole dataset. The voltage fluctuations above 200 microvolts per 250 ms and below 1 microvolts per 250 ms were excluded. The statistical analyses (parametric t-test for normal distributions) have been performed on the whole distribution of the data points calculated by FAA. The values at 95% significance level have been marked.

GSR Recordings and Data Analysis. Secondly, GSR method, as a measure of physiological arousal, has been used for exploring the changes in arousal dimension. The changes in arousal have already been issued with experience of difficulties while using an interface thus have been associated with frustration and stress. GSR recordings were performed via a mobile device (Shimmer) with a sampling rate 50 Hz. The fingers of the participants were cleaned with a soft cleaning gel. Then the device was plugged. The acquired data was filtered with a cut-off frequency of 0.3 Hz and with Butterworth filter of 10 Hz. The obtained continuous signal was detrended to eliminate the potential trend-related effects. The signal was then statistically analyzed to find the peaks throughout the whole distribution. Thus, the statistically significant peaks have been illustrated in a video of 250 ms frame (4 Hz). The obtained data has been statistically analyzed to find the significant peaks within the whole dataset. The values at 95% significance level have been marked and reported accordingly (Fig. 2).

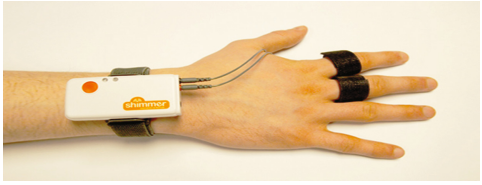

Fig. 2. Wireless Shimmer GSR device

Eye-tracking Records and Data Analysis. An eye-tracker device, Eye Tribe, has been used for data collection during the experiments. The eye-tracking data are important for understanding the trends in the EEG/FAA and GSR data. The eye-tracking data has been converted into heatmaps @4 Hz to be compatible with the other neurometric and biometric outputs.

4 Results and Discussion

4.1 Evaluation of the Findings from Conventional UX Research Methods

In this part, the findings obtained via different methods are summarized and demonstrated.

Evaluation of the Findings from the Task-Observation Stage. The product category menu on the left side appears as an initiator and navigator tool for the product search experience. Most of the participants (n = 14) attempted to make use of this menu especially in the very early periods of product search.

Search engine box, which provided autocomplete suggestions, was also used by most of the participants as a means for product search (n = 15). However, it was observed that only 4 participants preferred to use the search box as the first strategy to find the product they are looking for. Thus, as suggested by Baymard Institutes reports, the position, contrasts and size of the search field possibly affect the usage of this area. The remaining 11 participants preferred to search products through the category menu, but when they failed to find the desired product through the product category menu, they were inevitably forced to use the search engine box.

Some of the participants (n = 6) were also interested in the items that were presented in the main carousel in the categories. Specifically, a significant discount presented in the first slide of the carousel took their attention. Moreover, some of these participants made use of the navigator inside the carousel to change its content. However, the arrow buttons inside the carousel were not easily recognizable and appeared as a potential issue to be resolved.

One of the mostly observed behavior was that more than half of the participants (n = 11) were not able to notice the filter menu in the search results page due to the fact this menu is located below the category structure. Five participants could not see the filter menu during their entire product search process, whereas the remaining 6 participants could notice the filter menu far lately. This resulted in the insistent use of pagination and visiting the product search result screens one-by-one (n = 16).

One of the other aspects has been related to the use of filters. Only 9 participants directly head to the filter menu. Even by including the number of participants that noticed the filter menu late in their product search process, all the participants who used the filter menu (n = 15) limited their use with only three filters: Firsthand/ Second hand (*"Durum/Sıfır"*), free cargo (*"Kargo/Ücretsiz"*) and price range (*"Fiyat aralığı"*). Thus, our results on filter menu showed the importance of detailed and category specific filters as suggested by Holst (2016).

It was also remarkable that some of the participants (n = 4) were not able to see the arrow button to activate the price range filter after they entered the information to the

relevant boxes. These participants expected to have an automatic activation of the filters as they entered the price ranges. It was also observed that the complementary features to support product search such as "sorting" and "viewing options" were used by none of the participants. Only one participant attempted to use the "viewing options" to see the product search results in a row based layout. This finding appeared as a potential limitation for an effective use of it by most of the participants.

Evaluation of the Findings from In-depth Interview Stage. The findings from the in-depth interviews are in line with the findings derived from the task-observation stage. A summary of the findings, which are accompanied by relevant quotes from the users, are presented in Table 1 below. The findings fell into five major categories: Browsing with the use of left-hand category menu, Search experience, Use of persuasive design features, Faceted-Browsing (Use of Filters), Product Listings (Product Search Results Page).

Table 1. Summary of the findings from the in-depth interview stage

	Positive insights	Frequency	Quotes
Browsing with the use of left-hand category menu	• Use of the product category menu on the left side as an initiator and navigator tool for the product search experience	11	• "I can easily find what I'm looking for according to product groups." (K1) • "It is very well classified, everything what I'm looking for is present… The menus on the left side are explanatory and detailed. Inside of the menu (mega drop-down), presence of photos during access of sub-categories is good." (K11)
Search experience	• Use of the search engine box which is supported by autocomplete suggestions	5	• "I can find products easily. Google search, let's think about such a culture. I wrote 'kol', consequently, everything related to "kol saati" was appeared. I liked it." (K5)
Use of persuasive design features	• Presenting discounted items in the first slide of the carousel	5	• "I view the carousels like this. There have been always very nice goods in that place." (K7) • "It is being today's deal and 40% discount attracted my attention. The price dropped to 259 TL from 429 TL." (K3)

(*continued*)

Table 1. (*continued*)

	Positive insights	Frequency	Quotes
Faceted-Browsing (Use of Filters)	• When noticed, although three filters were used frequently, the filters were valued	9	• "I like choosing search criteria (filters). The field that we can search, either fill it with special words or enter any criterion. The product that I intended to buy… Is it new? Is it in its own box? Even these are important for me." (K13)
	Negative insights	Frequency	Quotes
Faceted-Browsing (Use of Filters)	• Inability to notice the filter menu in the search results page due to the fact that this menu is placed/presented below the category structure	11	• "I could not find filtration. It was not clear." (K1) • "The website is nice but with few details, I could not find many goods that I am looking for, I intended to enter price range but I could not find." K6 • "Filter menu is shaded in color, it does not attract attention." (K12)
	• Inability to notice the filter menu resulted in the insistent use of pagination during the navigation through product search result screens	14	• "If I had been able to see the price range, I would not have gone until the page 24" (K11)
	• Frustration caused by the lack of auto-filtering which resulted in the ignorance of the filter activation button	4	• "When we click on checkboxes how it renews the process, I thought it will do it for price range, too. But, it did not. I realized the arrow below afterwards." (K7) • "Because I could not see the confirmation button "OK" after I enter the price range during filtering, I could not click on it. It should be clearer. Dark blue, green, etc." (K1)

(*continued*)

Table 1. (*continued*)

	Negative insights	Frequency	Quotes
	• Lack of category-specific filters	4	• "I did not like the detailing in product search… I am looking for bicycle but I could not make a detailed search. Disc, brake, aluminum case. There is not enough criterion." (K4) • "The shoe size does not appear as a filter." (K8)
Product Listings (Product Search Results Page)	• Lack of detailed product information in product search results page	9	• "Although product names are the same, their prices are different. This got me confused. Also, I hesitated whether it is a second-hand product or not." (K6) • "I cannot see whether the product that I have bought has a guarantee or not." (K6)

The evaluation of the in-depth interviews for the product search task indicates that the participants were generally satisfied with the product search features. For instance, more than half of the participants (n = 11) were content in using the extensive product category menu on the left side as a tool to start their the product search experience.

As suggested by Appleseed (2014), the autocomplete suggestions feature in the search engine was expressed as a user-friendly and effective feature of this e-commerce site by some of the participants (n = 5). However, it should be mentioned that these participants were the ones which started their search process directly from the search engine box.

Presentation of discounted items in the first slide of the carousel were appreciated by some of the participants (n = 5). These participants mentioned that they tended to look at the prices first and then they were engaged in the product features. This might be related to one's price-sensitive position during e-shopping.

On the other hand, the participants (n = 11) could not notice the filter menu in the search results page due to the fact the filters were located below the category structure that was a difficult place to notice. Five of the participants told that they were even unaware of such a menu. Thus, the recognition of this menu in the product search results page appeared as an important problem. The participants who noticed and use the filter menu were content in using such a feature (n = 9). However, the participants who could notice the filter menu could only make limited use of it. Only three filters were frequently used by the participants: Firsthand/ Second hand (*"Durum/Sıfır"*), free cargo (*"Kargo/Ücretsiz"*) and price range (*"Fiyat aralığı"*). Only 4 participants declared the need for more category-specific filters.

Four participants were frustrated since they expected to find the automatic activation of the price range filter rather than the obligation to click on an action button. This finding supports the notion that automated filtering in faceted search is an important expectation among e-commerce users.

Findings from the in-depth interviews confirmed the observations on the insistent use of pagination by most of the participants (n = 14). The findings revealed that the inability of the participants to notice the filter menu in their product search process resulted in the insistent use of pagination during the navigation through product search result screens. They spent great efforts to visit the product search result screens one-by-one. This factor might be the main cause of the high amount of durations devoted to the product search stage. This observation could also be interpreted as part of a potential window shopper behavior. However, the findings from the in-depth interviews showed that only three of the participants associated their use of pagination with window shopping.

Almost half of the participants expressed difficulty in scanning through the product search results page due to the limited amount of product information presented in each search result (n = 9).

Evaluation of the Findings from the Surveys. Findings from the surveys overlapped with the ones derived from the previously adopted qualitative methods. The survey was composed of three main aspects: Perceived ease-of-use, satisfaction, and perceived time. The participants were asked to rate their experience during product search with respect to these aspects. The ratings ranged from 1 (lowest) to 5 (highest). The results are Perceived ease-of-use: 3.95/5, Satisfaction: 3.8/5, and Perceived Time: 3.45/5.

Perceived ease-of-use. The obtained results from the surveys indicate that the perceived easiness of the use of internet site for product search is generally positive (significantly higher than average). The participants' introspective evaluations generally focus on the fact that they could find the target products easily. Only 3 participants (P4, P13, and P21) responded lower than the middle rating as all of them had difficulty in finding the products they wanted to buy.

Satisfaction. The ratings of the participants for the satisfaction dimension during product search is also significantly higher than the middle rating (3). The ratings of the 3 participants, who spent considerable amount of time in finding the products they wanted to buy, were quite lower for the satisfaction index.

Perceived Time. The ratings for this index reflects that the participants completed their task of product search almost on average. The mean of ratings was not significantly different from the middle rating (3). Similarly, the ratings from the three participants was below the average.

It was also of our interest to observe if there is significant correlation between these three aspects. The results of the statistical analyses indicate that these aspects are significantly correlated with each other (dep. samples t-test; $p < .05$). Thus, the participants were generally tended to evaluate their satisfaction levels in relation to the perceived ease-of-use and perceived time.

4.2 Findings from the Neurophysiological and Psychophysics Methods

The findings from EEG-GSR measures are presented in this part. As mentioned above, the measures are used to observe negative experiences of the participants relying on FAA and peaks in the GSR level (Table 2).

Table 2. The appearances of the negative signals are marked with "1" and "0" sign indicates that there is no such observed negative peak for the given activity. (P3 was excluded because of the synchronization problems during data collection)

	P1	P2	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	No	%
Use of search filters	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	17	89
Browsing experience (of the prodcut categories)	1	1	0	0	1	1	1	0	0	0	0	1	1	1	0	0	1	0	0	9	47
Excessive page loading duration	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	11
When unable to find a button	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	11
Scrolling up & down	1	0	0	0	1	1	1	1	0	1	0	0	1	0	0	0	0	0	0	7	37

The most common neural pattern has been related to the use of search filters. It might have possibly been due to the fact the participants experienced problems while using the filters. The obtained findings indicate that the use of filters appears as a potential problem for almost all the participants (79%). The negative FAA peaks during the attempts for the filters (such as price range) (Fig. 3).

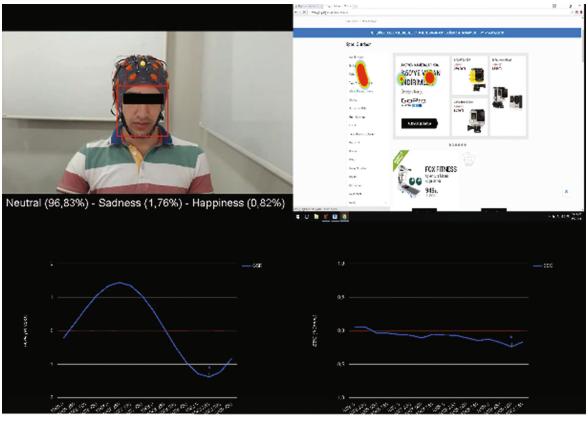

Fig. 3. The participant is experiencing negativity while finding the relevant product category during product search phase.

Moreover, it is also remarkable that when the participants were not able to find the location of certain buttons of interest (such as "add to cart" or "sorting option"), the negative FAA peaks were observed and reported. These might be assessed as problems due to usability, since user experience (UX) has classified to be negative.

Moreover, scrolling down and up with the use of mouse has been observed to be a potential precursor of negative affect that might be because of the difficulty of tracking the flow on the screen (n = 7). The issue related to the situations that the participants could not find the relevant button on the webpage was infrequent (n = 2), however, these cases resulted with clear incidences of negative affect.

Results of the Time-on-Task (TOT) Measures. The completion time for product search task was 754.047 s. The analysis of the Reaction Time data indicates that the individual differences have a considerable impact on the high level of variance. This fact could be explained with the differences regarding the participants' intentions related to purchasing. Some could easily determine the target product and find suitable alternatives quickly. However, the rest experienced difficulties while deciding on the product category. As initially expected, the standard deviation of this measure for this task was higher when compared to the average ToT measures.

4.3 Comparison of the Findings from Multimodal Methods

The findings from neurophysiological and psychophysics methods indicate that there are number of UX issues that are in line with the findings from different methods. The most common ones are about the issues related to the inefficient use of filters validated by both methods. The use of filters under the product category menu has been limited that have been reported by task-observation and in-depth interview methods. The findings from neurophysiological methods indicate that 79% of the participants experienced negative affect while using these filters. Moreover, 3 out of 4 participants who have complained about the insufficiency of category-specific filters (during the in-depth interviews) are reported to have negative affect during the use of filters. Thus, the observed negative affect might be because of their insufficient level of satisfaction during product search and specifically related to the ineffective use of these search filters. The table below shows the interconnection across the findings between different methodologies (Fig. 4).

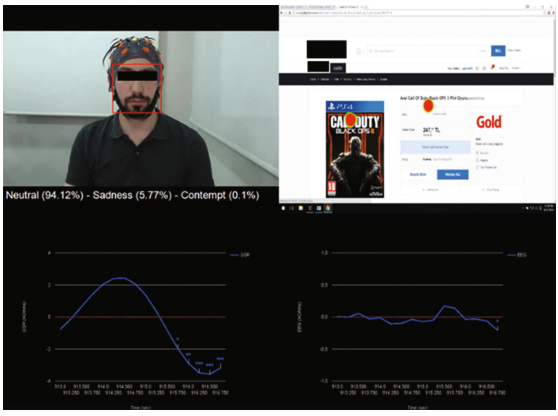

Fig. 4. The participant is experiencing negative affect because of the unexpected product price

The findings obtained from neurophysiological/psychophysics methods indicate that two of the participants were likely to experience negative affect as the page loadings reached excessive durations. On the other hand, scrolling up and down through the page were observed to cause negative affect for some of the participants (7 out of 19 participants). Moreover, two of the participants were observed to have negative affect, since they could not find the relevant buttons on the page. This might have caused a considerable level of anxiety for these participants. But it is important to remark that these neurophysiological/psychophysics findings were not supported by the qualitative methods. These might be accepted as contributive findings by these quantitative methods. However, a better methodology might be to ask the participants about this kind of potential cases (Table 3).

Table 3. This table shows the findings related to the use of filters by neurophysiological/psychophysics, in-depth interview and task observation methods

	1	2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	No
Negative affect by Neuro/Bio methods during use of search filters	1	1	1		1	1	1	1	1	1		1	1	1	1	1	1	1	1	17
Frustration caused by the lack of auto-filtering which resulted in the ignorance of the filter activation button	1					1	1				1									**4**
Inability to notice the filter menu resulted in the insistent use of pagination during the navigation through product search result screens	1	**1**					1	1		1	1	**1**	1		1	**1**	1	1	1	**13**
Lack of category-specific filters			1					1	1							1				**4**
Inability to notice the filter menu in the search results page due to the fact this menu is placed below the category structure	1			**1**			1	1		1		**1**	1		1	1			1	**10**

(*continued*)

Table 3. (*continued*)

	1	2	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	No
Inability to notice the filter menu in the search results page due to the fact this menu is placed below the category structure	1			1			1	1		1		1	1			1				**8**
Expectancy of auto-filtering which resulted in the ignorance of the filter activation button	1					1	1													**3**

It is also important to remark that the individual differences might be influential to a certain extent. For instance, one participant might be negatively affected because of waiting for the load of a new page whereas the other might not have a problem with this. Thus, overall evaluations and inter-methodological links have such a potential problem while trying to find the relevant explanations regarding the given context. One future solution might be the use of a deep-dive survey that investigate most of the potential problems for each of the participant. Hereby, it will be possible to provide a useful framework for considering the individual differences just after the user has experienced it and asked specifically about it (Fig. 5).

Fig. 5. The participant is experiencing negative affect due to the fact product information page does not satisfy him during product search phase

In relation to the very early discussions about the relative positions of the newer and older methods, it is necessary to provide an insight for the current state-of-art. The obtained findings from multi-modal methods suggest that the neuroscientific/psychophysics versus conventional methods might come up with partially overlapping but in general diverging results as the outputs of the current study indicates. On one hand, it is quite convincing if both the subjective and objective methods end up with converging results as in the use of search filters. However, this is not the case for several occasions and the outputs from these methods do not seem to overlap although they do not contradict each other. In other words, these outputs might be focusing on different aspects of the user experience. It is quite trivial that the subjective methods are sensitive for the high-order cognitive processing that have been accounted by the user herself. On the other hand, the objective methods might be more sensitive for detecting lower level affects that are not generally articulated by the participant. Thus, the main contrast might be linked to the ongoing debate between explicit and implicit levels of information processing. Hence, the outputs of these different methodologies might be used as complementary tools for each other.

5 Conclusion

Overall, the obtained empirical findings of this exploratory study enabled to point out successful features to support a seamless buying experience in the target website. However, the findings also revealed a couple of user experience problems related to the design and implementation of the target website. The main issues obtained by the qualitative methods including task observations and in-depth interviews cover the search experience, browsing experience (through the left-hand category menu), use of filters and sorting, and product listing (product search results page).

On the other hand, the findings by neurophysiological/psychophysics methods indicate that the participants tend to have negative affect when they make use of search filters, browsing experience via category menu, scrolling up/down, and less frequently while waiting for page loading and when cannot find the searched button on a page.

Addressing these problems accurately are quite significant especially for enhancing the usability of the website. In this context, the findings of the current study are used by the UX designers of the website to improve the relevant experience.

There have been occasional findings from the neuroscientific/psychophysiological methods and the conventional methods that seem to support each other to a certain extent such that the use of filters have been found to be problematic.

However, there are also diverging findings such that qualitative methods indicate that the filter menu in the product search page is quite unrecognizable.

On the other hand, neurophysiological/psychophysics findings indicate that the scrolling up/down, not being able to find the relevant buttons on the page, and excessive page loading durations have also been observed as potential precursors of negative affect.

Lastly, there are findings for which qualitative and neurophysiological/psychophysics methods contradict each other such as the findings related to the browsing a product from the category menu.

Despite the fact the potential use of neurophysiological/psychophysics methods is still controversial and needs many more direct research and valid findings in this specific area, it is also arguable that such multimodal use will provide fruitful contributions with the integrated use of conventional and neuroscientific methods.

Acknowledgement. This work was financially supported by gittigidiyor (eBay). We'd like to thank for the efforts and support of Çağrı Karahan and Oğuzhan Poyrazoğlu in the realization of this study.

References

ISO DIS, I. 9241-210: 2010. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems. International Standardization Organization (ISO), Switzerland (2009)

Chai, J., Ge, Y., Liu, Y., Li, W., Zhou, L., Yao, L., Sun, X.: Application of frontal EEG asymmetry to user experience research. In: Harris, D. (ed.) EPCE 2014. LNCS, vol. 8532, pp. 234–243. Springer, Cham (2014). doi:10.1007/978-3-319-07515-0_24

Hornbæk, K.: Current practice in measuring usability: challenges to usability studies and research. Int. J. Hum. Comput. Stud. **64**(2), 79–102 (2006)

Drenner, S., Sen, S., Terveen, L.: Crafting the initial user experience to achieve community goals. In: Proceedings of the 2008 ACM Conference on Recommender Systems, pp. 187–194. ACM, October 2008

Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P., Kort, J.: Understanding, scoping and defining user experience: a survey approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 719–728. ACM, April 2009

Boehner, K., DePaula, R., Dourish, P., Sengers, P.: Affect: from information to interaction. In: Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility, pp. 59–68. ACM, August 2005

Ravaja, N.: Contributions of psychophysiology to media research: review and recommendations. Media Psychol. **6**(2), 193–235 (2004)

Briesemeister, B.B., Tamm, S., Heine, A., Jacobs, A.M.: Approach the good, withdraw from the bad—a review on frontal alpha asymmetry measures in applied psychological research. Psychology **4**(03), 261 (2013)

Davidson, R.J., Schwartz, G.E., Saron, C., Bennett, J., Goleman, D.: Frontal versus parietal EEG asymmetry during positive and negative affect. Psychophysiology **16**, 202–203 (1979)

Ohme, R., Reykowska, D., Wiener, D., Choromanska, A.: Analysis of neurophysiological reactions to advertising stimuli by means of EEG and galvanic skin response measures. J. Neurosci. Psychol. Econ. **2**(1), 21 (2009)

Tomico, O., Mizutani, N., Levy, P., Yokoi, T., Cho, Y., Yamanaka, T.: Kansei physiological measurements and constructivist psychological explorations for approaching user subjective experience. In: DS 48: Proceedings of the 10th International Design Conference, DESIGN 2008, Dubrovnik, Croatia (2008)

Wheeler, R.E., Davidson, R.J., Tomarken, A.J.: Frontal brain asymmetry and emotional reactivity: A biological substrate of affective style. Psychophysiology **30**(1), 82–89 (1993)

Boucsein, W.: Electrodermal Activity, p. 2. Springer Science & Business Media, New York (2012)

Westerink, J.H., Van Den Broek, E.L., Schut, M.H., Van Herk, J., Tuinenbreijer, K.: Computing emotion awareness through galvanic skin response and facial electromyography. In: Westerink, J.H.D.M., Ouwerkerk, M., Overbeek, T.J.M., Frank Pasveer, W., de Ruyter, B. (eds.) Probing Experience. Philips Research, vol. 8, pp. 149–162. Springer, Netherlands (2008)

Carlson, N.R.: Physiology of Behavior. Pearson, Boston (2013)

Ohme, R., Reykowska, D., Wiener, D., Choromanska, A.: Application of frontal EEG asymmetry to advertising research. J. Econ. Psychol. **31**(5), 785–793 (2010)

Holst, C.: Consider 'Promoting' Important Product Filters (80% Don't), 5 April 2016. http://baymard.com/blog/promoting-product-filters. Accessed

Appleseed, J.: 8 Design Patterns for Autocomplete Suggestions, 1 July 2014. http://baymard.com/blog/autocomplete-design. Accessed