User Interface Evaluation of Interactive TV: A Media Studies Perspective[*][+]

KONSTANTINOS CHORIANOPOULOS[1] AND DIOMIDIS SPINELLIS[2]

[1]*Imperial College London, Department of Electrical and Electronic Engineering, London SW7 2BT, UK*

[2]*Athens University of Economics and Business, Department of Management Science and Technology, Patision 76, GR-104 34 Athens, Greece*

Email: k.chorian@imperial.ac.uk, dds@aueb.gr

Abstract

A diverse user population employs interactive TV (ITV) applications in a leisure context for entertainment purposes. The traditional user interface evaluation paradigm involving efficiency and task completion may not be adequate for the assessment of such applications. In this paper, we argue that unless ITV applications are evaluated with consideration for the ordinary TV viewer, they are going to be appropriate only for the computer literate user, thus excluding the TV audience from easy access to information society services. The field of media studies has accumulated an extensive theory of TV and associated methods. We applied the corresponding findings in the domain of ITV to examine how the universal access to ITV applications can be obtained. By combining these results with emerging affective quality theories for interactive products, we propose a user interface (UI) evaluation framework for ITV applications.

Keywords: interactive television, user interface, affective quality, media studies, evaluation, methodology

1 Introduction

Despite the rapid growth and wide adoption of the personal computer (PC), the Internet and the mobile phone, TV remains the most popular and the most widespread electronic medium. The diffusion of TV in some countries is reaching the majority of the households, while TV watching consumes the largest share of leisure time [52]. Lately, the transition towards Digital TV (DTV) has transferred the information technology features of the PC and the Internet into the digital set-top box (STB), which is used to introduce various interactive applications through the TV set. It has been argued that the promise of interactive services for all members of the society may remain unfulfilled, unless the usability of the new medium is adapted to the diverse characteristics of the population [14] [41]. In addition, Monk [28] argued that there is a need to adapt the traditional UI design and evaluation methods to the home environment. Since ITV applications serve entertainment goals and domestic leisure activities for a diverse user

---

[*] Konstantinos Chorianopoulos and Diomidis Spinellis. User Interface Evaluation of Interactive TV: A Media Studies Perspective. *Universal Access in the Information Society*, 5(2):209–218, August 2006.

population [23], there is a need to re-examine the traditional usability engineering concepts and evaluation methods, under the light of existing results from the field of media studies. Indeed, the intersection between human-computer interaction (HCI) and the mass communication disciplines has been highlighted as a significant area for further research [25]. In this paper, we argue that universal access can easily be extended with the proposed methods and techniques, since it is —by definition— receptive to the diversity in the user population, application domain, and context of use.

Most previous HCI research about ITV has focused on the design of the Electronic Program Guide (EPG), which can be described as an information retrieval and navigation system. Accordingly, the majority of previous evaluations of ITV applications have employed an efficiency conceptualization of the UI, without any consideration of the entertainment needs of TV viewers. At the same time, the affective dimension of the UI is gaining ground in contemporary HCI research. Previous work has addressed the affective quality of Web sites [22], automatic teller machines [44], PC applications [17] and ITV applications [3], but currently there is no methodological framework for defining and evaluating the quality of an ITV UI. For this purpose, we integrated research from affective HCI with media studies, in order to devise a conceptualization for UI evaluation that facilitates universal access to ITV applications. In addition, we employed the three-level model of affect [33] and used it to organize a collection of relevant constructs and measuring instruments for ITV applications.

The findings of media studies and advertising research relate to the present work in various ways. Mass communication has explored the effects of broadcast electronic media messages to the TV audience. It has developed several important concepts, such as the "uses and gratifications" theory [39], which describes the motivations for watching TV. Such theory does not assume an attentive user like the traditional usability engineering methods do, but measures explicitly a continuum of viewer involvement with a TV program [36]. Moreover, the "selective exposure" paradigm [53] regards the viewer as an active receiver of the media messages, who changes TV channels and actively selects TV contents to be exposed to. The selective exposure concept contrasts with the traditional usability conception of a specific task to be performed by a user. Overall, we collected a relevant set of methods and concepts and used them to guide the development of an affective UI evaluation methodology for ITV applications.

The rest of this paper is structured as follows. In the next section, we establish the need to re-examine the traditional usability engineering conceptualization in the context of ITV applications, we present the shortcomings of previous usability evaluation studies for ITV applications and we give definitions of the affective quality of the UI. In Section 3, we develop an affective conceptualization of the UI for ITV applications, which is based on theories of mass communication and of affective quality. In Section 4, we discuss the suitability of alternative UI evaluation techniques. In Section 5, we present a set of constructs and instruments, which are appropriate for the evaluation of ITV UIs. In section 6, we put the usability framework for ITV in the broader context of HCI, universal access and ITV research. Finally, we give recommendations for further research.

2 Universal Access and Interactive TV

To find the issues that should be addressed in the evaluation methodology of an ITV UI, we will review research findings from: 1) usability engineering and universal access, 2) affective quality theories, and 3) ITV evaluation studies.

From Usability Engineering to Universal Access

Usability is defined as ease of use, and it is associated with the efficient use of an interactive product [31]. One basic element in the conceptualization of usability is the notion of user task. A user task consists of a finite number of steps and has an exact ending. Accordingly, a usability evaluation session includes a few tasks that should be performed by a user. The objectives of a usability evaluation session have been classified into five broad categories: 1) learnability, 2) memorability, 3) efficiency, 4) effectiveness and 5) user satisfaction [31]. Each one of the above objectives could be deconstructed into more concrete and measurable goals. For example, effectiveness is inferred by counting the number of errors, or the number of successful task completions, while efficiency is inferred by measuring the time required to complete a task. User satisfaction is usually measured by eliciting the users' opinions about specific issues and may also include the subjective (perceived) usefulness and ease of use.

Previous research has established that the perceived ease-of-use and the perceived usefulness correlate positively with the user acceptance of a new technology [5]. The objective and the subjective usability measures are positively correlated [32], which means that a UI that is enabling fast task completion is also perceived as more usable. Accordingly, most research in the domain of PC software productivity and of Web usability has focused on improving the user performance. Nevertheless, more recent research has addressed the limitations of the above correlations and suggested that the application domain, the user's experience, and the context of use should also be justified [13]. In brief, the traditional usability engineering methodology is appropriate for the office environment, but the diffusion of computers targeting entertainment outside the typical office context has created the need for alternative conceptions.

Universal access is often seen as providing everybody with the means to get information and to perform tasks within a reasonable timespan and with a reasonable amount of effort. Compared to the traditional usability definition, universal access emphasizes the diversity in the user population, in the application domain and in the context of use [41]. Universal access methods have been applied to facilitate the accessibility of information society services for the disabled, aged, and children. Interactive TV is an information society technology that is employed in a leisure context of use, is targeted to the majority of the population, and might provide a terminal for diverse activities, such as e-commerce, e-learning, and games. At first sight, contemporary universal access techniques seem to be appropriate for the evaluation of ITV applications. Still, the TV audience has been accustomed to expect much more than ease of use. In particular, the TV audience receives information and expects to be entertained, in a layed-back posture and through an emotionally loaded visual language. The universal access paradigm has not yet considered this affective dimension of the interactive TV experience, and thus lacks the methods and techniques to serve the needs of ITV users. It could be argued that the current definition of universal access is not sufficient and that we need to regard universal access as an inclusive affective experience. In this way, having satisfied the basic usability requirement, everybody should be receiving a reasonable level of entertainment.

Affective Quality of Interactive Products

The explicit distinction between the hedonic and the utilitarian experience is anything but novel in the academic literature. Previous works in consumer research have addressed the differences between information processing and experiential consumption [19]. Similarly, in HCI research, there is growing evidence that the traditional desktop usability concepts do not account for the pleasure of the user experience [17] [45]. In particular, Tractinsky [44] found that the perceived ease of use correlated positively with the aesthetics, but not with the actual efficiency in task completion. In other words, the usability of the most aesthetically pleasing UI was considered better than the usability of the most efficient one.

In recent years, an increasing number of HCI researchers have investigated the role of emotions in the design and evaluation of a UI [7] [33]. Contemporary research [16] [51] suggests that the quality of interactive products consists of three elements: utility (usefulness), ease of use (usability), and enjoyment (affective quality). These elements should be regarded independently, although there is evidence that perceived usability depends on affective quality [44] [45]. Anyhow, the affective quality is very relevant for the UI evaluation of ITV applications, and the respective methodology could benefit by applying concepts of media studies, the established field of TV research.

The Case of ITV Applications

The evaluation of ITV UIs is an important factor for the adoption of new services [26]. HCI researchers have recognized the significance of this emerging application domain and have performed numerous evaluations of ITV UIs. Previous usability tests of the EPG and the video navigation [1] [8] [10] [49] have employed the traditional usability engineering concepts, such as task efficiency and effectiveness. Indeed, EPG usability aspects are very similar to those of productivity software, because the interface involves more information processing than enjoyment of ITV content. Several aspects of EPG navigation can be modelled after traditional HCI tasks and goals. Nevertheless, there are some aspects of EPG design, and many other types of ITV applications that would benefit by a consideration of the affective dimension of the UI.

Most notable among the recent findings for ITV applications is the realization that users' subjective satisfaction is at odds with the established metrics of efficiency. For example, a usability test of three video skipping UIs revealed that user satisfaction was higher for the UI that required more time, more clicks and had the highest error rate [8]. In other words, the most efficient UI was not the most favoured one. This result contravenes the assumptions of the efficiency usability paradigm, which conceives efficiency as equivalent of usable and satisfying. The satisfaction questionnaires exposed that the users regarded their preferred UI as more fun and relaxing compared to the most efficient one [8]. Accordingly, in another experiment, we let our subjects use a video skipping application without specifying any task, besides the suggestion to 'watch TV for a period of time'. Moreover, we employed the hedonic quality construct [17], and the results confirmed that users liked a video skipping UI, although it was coupled with a dynamic advertisement insertion feature that increased the total number of advertisements shown.

Summary of Issues and Approach

The majority of previous studies of ITV applications have considered only the efficient aspect of the UI. Because ITV applications serve entertainment aspirations in a leisure context and for a wide diversity of users, there is a need to extend the universal access toolset so that it considers the affective quality of an ITV UI. Table 1 summarizes the relationship between usability and affective qualities for four different UI evaluation dimensions.

Table 1. Methodological issues in the evaluation of ITV UIs

UI evaluation	Usability	Affective quality
User, task, context	PC user, productivity, work	Viewer, entertainment, leisure
Concepts	Task, effectiveness, efficiency	Entertainment, relaxation
Procedure	Task execution	Free exploration
Constructs	Task completion, errors, efficiency	Affective state, emotions

Existing work in media studies and advertising research presents techniques for measuring emotional responses to TV content. Furthermore, previous work in the HCI domain has addressed techniques for assessing the affective quality of a UI. In the following sections, we combine the latter two, to develop a UI evaluation framework for ITV applications (Figure 1).
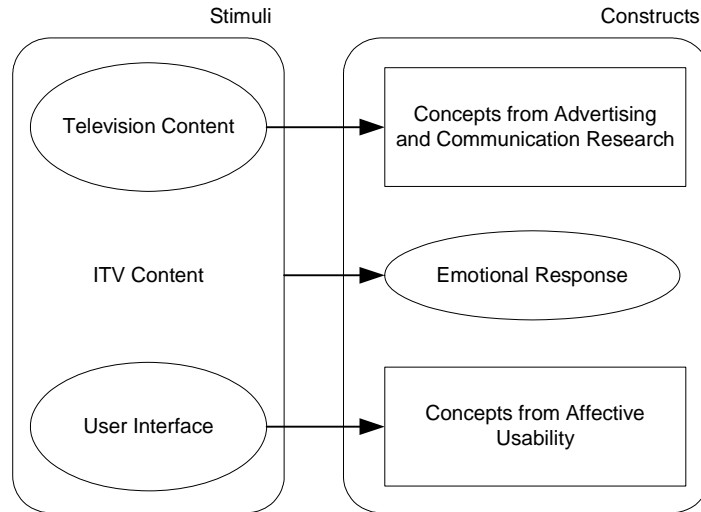

Figure 1 Emotional response to the ITV UI may be assessed either with the affective UI, or with the TV content techniques

3   The Quality of Interactive TV User Interfaces

It has been argued that people spend most of their leisure time trying to moderate their moods. Daniel Goleman [15] writes: "managing our emotions is something of a full-time job: much of what we do —especially in our free time— is an attempt to manage mood. Everything from reading a novel or watching television to the activities and the companions we choose can be a way to make ourselves feel better. The art of soothing ourselves is a fundamental life skill." Thus, television entertainment could be conceptualized as mood management [46]. Actually, television entertainment is a multidimensional construct that cannot be measured as a whole, but consists of several parameters that could be measured [47]. For example, Reeves and Nass [38] assert that a mediated experience elicits an emotional response, which is partly valence (pleasure) and partly arousal. There are also additional elaborate models of the uses and gratifications when watching TV [23] [39]. Therefore, the UI of an ITV application could be used as an additional —to channel changing and program selection— means to moderate the mood of the TV viewer.

In this paper, it is assumed that a user controls an ITV application for the purpose of regulating mood. Then, the evaluation could exploit those constructs that have been employed to assess viewers' emotional responses to TV content. Still, there is also a need to consider the interactive part of ITV, for which there is no extensive research in the mass communication discipline [46]. For this purpose, we employ concepts for measuring emotional responses to a UI from the area of affective UI research. Overall, measuring the ITV entertainment could be organized into two parts: 1) measuring emotional responses to TV content and 2) measuring emotional responses to a UI (Figure 1). The decomposition of the ITV UI evaluation into two parts is not meant to measure independently the parts. In other words, the decomposition of ITV in the UI and the TV component is merely an operational arrangement, which is employed to structure the presentation of the related research. Indeed, we argue that an ITV UI should ideally be designed to be an integral part of the audiovisual content. Next, we consider a classification for emotional responses.

Brain Structure

Reflective
Behavioral
Visceral

Emotional Response

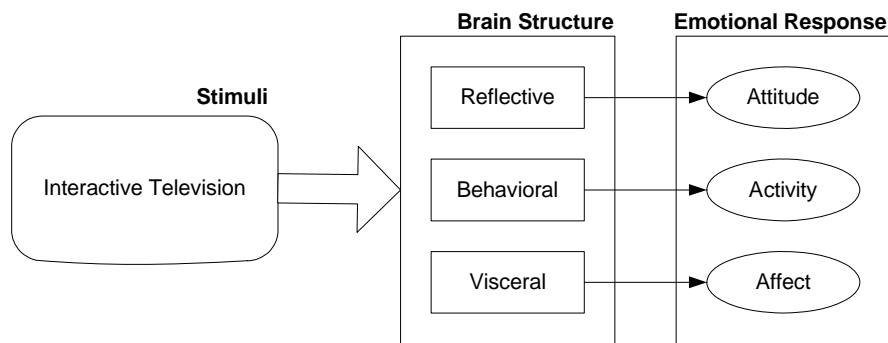Attitude
Activity
Affect

Stimuli

Interactive Television

Figure 2. The ITV entertainment experience elicits three types of emotional responses (attitude, activity, affect), which correspond to the three level model of affect —adapted from Norman et al. (2004)

According to Norman [33], there are three distinct levels of brain mechanism: 1) the visceral level, which is the pre-wired part of the brain and acts automatically to external stimuli, 2) the behavioural level, which contains the brain processes that control everyday behaviour and 3) the reflective level, which is the contemplative part of the brain. Each level could be associated to a different class of constructs, which could then be employed to evaluate the differences between the emotional responses to alternative UI designs (Figure 2). For example, an ITV application may elicit enjoyment (e.g., pleasure, or arousal) at the visceral level. Then, the user may continue using the ITV application for a long time and become emotionally absorbed (e.g., involvement and engagement). Finally, the user may decide that she likes the specific ITV application, which leads to the formation of an attitude (e.g., program liking).

Before presenting a collection of constructs and the respective measuring instruments, in the next section, we discuss UI evaluation methods and techniques for ITV applications.

Evaluation Methods

The choice of a specific UI evaluation method depends on the type of research problem to be addressed. For example, an ethnographic study may provide in-depth insights about the uses of TV in everyday domestic life [34]. Then, a survey may reveal relationships between the uses and the type of the family or the viewers' profile, and to provide quantitative results [12]. Previous findings demonstrate that the consumers' perceptions and especially the mental models that they form for new domestic technologies are very elastic and prone to change with time [37]. Therefore, a longitudinal study could also be used to study the evolution of important variables for longer periods of time [21]. In addition, ITV UI research has employed focus groups and interviews [9]. The latter methods are useful for requirements collection and for investigating the long-term effects of ITV applications, while usability tests are more suitable during the development process.

The majority of UI evaluation studies have been conducted in the laboratory with experimental methods. Mass communication research employs large (compared to HCI experiments) samples of people, in order to study the effects of TV content on viewers. On the other hand, HCI research focuses on informing product development and employs small numbers of subjects iteratively with discount usability engineering techniques [31]. Maguire [26] raised the research question of whether tasks should be fixed, or users should be allowed to use the service as freely as they wish. It has been argued that the users should be allowed to use the service for a predefined, but flexible duration of time (e.g., 15–30 minutes), without any particular task to complete [3]. In this way, the traditional usability test reflects the tradition of the selective exposure paradigm [53], which has been also used to study the media effects of interactive products in contemporary mass communication research [20]. Because viewers select TV channels

and watch TV programs in order to regulate their mood, the evaluation of an ITV UI should facilitate free exploration and enjoyment of the ITV application.

Every evaluation method employs one or more qualitative or quantitative data collection techniques. Many evaluation studies employ qualitative techniques (e.g., observation, thinking aloud, interview, focus group), but some techniques (e.g., thinking aloud) may not be suitable for ITV [26]. Quantitative techniques provide explicit results for formulated hypotheses and concrete UI issues, while qualitative methods are used to reveal UI issues that have not been identified by the designers. Ideally, the qualitative measurement techniques should be used to complement the quantitative ones [9].

In the next section, we focus on quantitative measurement techniques, which could be employed in affective UI evaluation.

4 Data Collection Techniques

Previous research has developed a number of techniques for measuring emotion, which range from physiological measures to iconographic scales [6]. The emotional response at the visceral level (Figure 2) may be inferred by a physiological measure (e.g., EEG, skin conductance, heart rate, and facial expressions), language, and behaviour, but the most popular technique for UI evaluation is self-report. The emotional response at the behavioural level may be detected by the analysis of the interactivity logs, and self-reports that convey the attention, involvement, and engagement of the user. Finally, the attitudes can be measured straightforwardly through retrospective questionnaires.

The rest of this section provides a critical overview of the constructs and the instruments that have been used to measure emotional responses to TV content or to a UI. The underlying objective is to develop a usability evaluation framework that is appropriate for measuring the affective dimension of a UI.

Emotional Responses in Media Studies and Advertising

A review of the literature in advertising and media studies revealed a number of concepts and their respective measurement instruments (Figure 3). The most relevant have been selected and are presented bellow: 1) Affect and activation, 2) involvement, and 3) program liking.
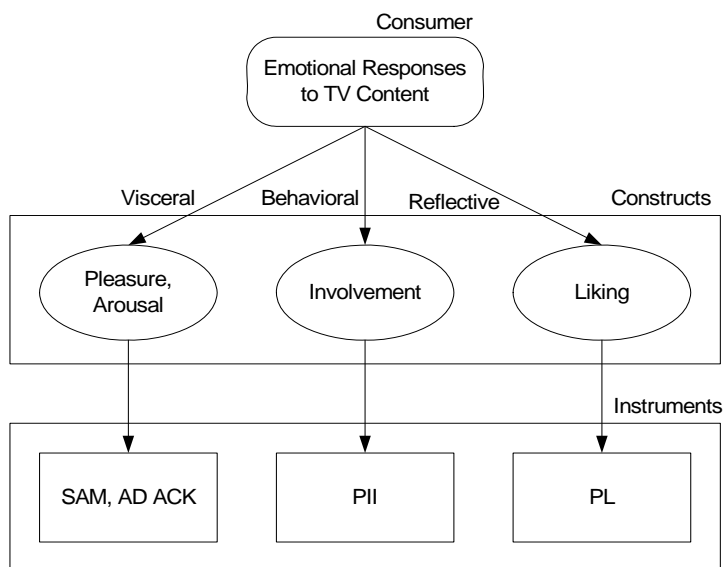

Figure 3. Three indicative constructs and the respective measuring instruments for evaluation of the emotional responses to TV content

Pleasure and Arousal

Most of the theories agree that there are three distinct dimensions of affect: 1) Pleasure (also called valence), 2) Arousal and 3) Dominance [40], called the PAD model of affect. According to the PAD model, all emotions can be accurately described in terms of three independent and bipolar dimensions: pleasure-displeasure, degree of arousal (runs from aroused to asleep), and dominance-submissiveness. These elements are autonomous as differing values along any of these three dimensions can occur concurrently without affecting each other.

These dimensions can be used to roughly describe the affective state of a person, although they do not give any information about the specific emotion being felt. The above constructs correspond to the visceral part of the brain (Figure 2), which is the source of the instinctive responses to external stimuli. One popular and easy to administer iconographic instrument for the PAD model is the Self Assessment Manikin (SAM) [2]. In the past, the SAM has been widely used in consumer and advertising research to record human affect for a variety of stimuli, such as mediated experiences, products, and service encounters. The SAM consists of three iconographic scales, each one corresponding to one of the three dimensions of the PAD model of affect (Pleasure, Arousal, Dominance). In this way, users indicate their emotional status by checking below or between the icons, along a 9-point scale.

Besides the PAD model, the Thayer's [42] Activation Deactivation Adjective Check-List (AD ACL) could be employed, which is a more elaborated model of a person's arousal, since it regards two possible dimensions of arousal. The subscale adjectives are as follows: Energetic (active, energetic, vigorous, lively, full-of-pep); Tired (sleepy, tired, drowsy, wide-awake, wakeful); Tension (jittery, intense, fearful, clutched-up, tense); Calmness (placid, calm, at-rest, still, quiet). The AD ACL employs a four-point self-rating system for each adjective in the list (Table 2).

Table 2. An explanation of the rating scale of Thayer's [42] Activation Deactivation Adjective Check-List (AD ACL)

relaxed: ✓✓ ✓ ? no	If you circle the double check (✓✓) it means that you definitely feel relaxed at the moment.
relaxed: ✓✓ ✓ ? no	If you circle the single check (✓)) it means that you feel slightly relaxed at the moment.
relaxed: ✓✓ ✓ ? no	If you circled the question mark (?) it means that the word does not apply or you cannot decide if you feel relaxed at the moment.
relaxed: ✓✓ ✓ ? no	If you circled the no it means that you are definitely not relaxed at the moment.

Involvement

According to Park and Young [35], most researchers agree that the level of involvement can be understood by the degree of the personal relevance or importance. The involvement construct may be refined for the assessment of particular ITV content. For example, Vorderer et al [48] evaluated an interactive story by employing questionnaires measuring the empathy towards the protagonist and the suspense. The involvement construct corresponds to the behavioural part of the brain structure, because the degree of personal relevance is formed while watching a TV program. Zaichkowsky [50] has developed a widely used scale that measures the involvement with products or advertisements, called the Personal Involvement Inventory (PII). The instrument consists of 20 semantic differential items, such as: 'means a lot to me/means nothing to me,' 'boring/interesting', and undesirable/desirable.' Mass communication research has devised alternative scales for measuring involvement with TV content, but those scales

are usually developed in an ad-hoc fashion to measure specific types of TV content, such as news [36] and storytelling [48]. Therefore, a UI may be evaluated for general (e.g., PII) or for specific (e.g., suspense) involvement.

Program Liking

Previous research has made a distinction between the feeling states and the program liking [30]. This distinction is consistent with the contemporary conceptualization of the brain structure, in the visceral, behavioral,and reflective parts. The Program Liking (PL) construct corresponds to the reflective part of the brain, because an attitude toward an experience is built after watching specific TV content and deliberately thinking about it. Murry et al. [30] assert that, in contrast to feeling states, program liking is a summary evaluation of the experience of viewing a television program. For example, viewers may enjoy a movie that elicits negative feelings, because they know that it is not actually true. Accordingly, the UI may be designed and evaluated in a way that enhances the emotion of fear. A generic instrument for measuring program liking consists of the following items on a 7-point scale [30]: 'I'm glad I had a chance to see this program,' 'I would never watch a rerun of this program on television,' 'I liked watching this program,' If I knew this program was going to be on television, I would look forward to watching it,' 'I disliked watching this program more than I do most other TV programs,' and 'There is something about this program that appeals to me.' The program-liking construct should be refined for the assessment of particular ITV content. For example, Vorderer et al [48] evaluated an interactive story by employing questionnaires that measured several aspects of movie liking that are specific to storytelling (e.g., "Overall, the movie was suspenseful").

Emotional Responses in User Interface Research

A review of the HCI literature revealed some constructs that are associated with emotional responses. The most relevant have been selected and are presented below, together with their respective measurement instruments (Figure 4): 1) feeling states, 2) engagement, and 3) hedonic quality.
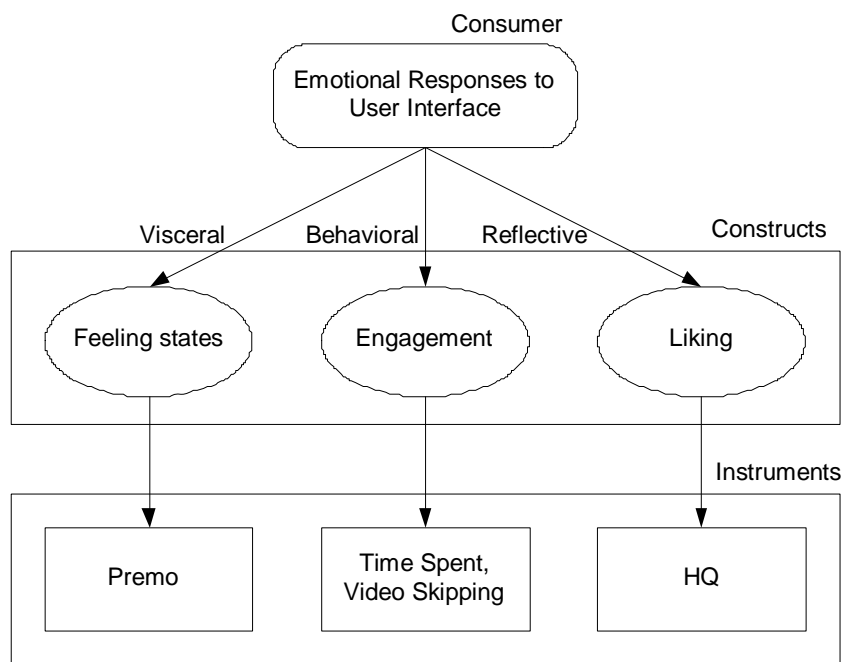

Figure 4. Three indicative constructs and the respective measuring instruments for the evaluation of emotional responses to a UI

Feeling States

Desmet [6] has developed an interactive animated pictorial questionnaire that is very similar to SAM, but it is not openly available (Desmet, personal communication). Therefore, the emotional response at the visceral level could be measured with the instruments from psychology research that were presented in the previous section.

Engagement

The involvement construct discussed in the previous section reflects the personal relevance and attention that a user pays to a mediated experience, but does not reveal the quantity of personal resources that the user is actually devoting to the ITV application. For example, a user may spend a few minutes attentively watching a broadcast TV programme, or a few hours of sparse use of an ITV application. For this reason, the engagement construct could be employed to capture how much interest is created by a mediated experience. The engagement construct corresponds to the behavioural level of the emotional brain model. Malone [27] measured the time spent using alternative UI manipulations, to get an insight about the interest of players in different versions of a video game. Correspondingly, if users spend more time with a specific ITV UI, it can be argued that it is more engaging. In addition, video skipping and other ITV navigation activities could be tracked to infer whether the user is attentive to the ITV application.

Hedonic Quality

After the users have interacted with a system for a certain period of time, the reflective level of the brain will be able to evaluate their affective state and their performance. As a consequence, an opinion will be formed regarding the appeal of the system. Indeed, studies in the field of affective UI evaluation have validated that users may form different opinions about the ergonomic and hedonic quality of a software product [18]. Previous research for a TV UI has included in the evaluation a single question, such as 'How much fun was the user interface' [8]. The hedonic quality construct corresponds to the reflective part of the brain structure, because it assumes a rational judgement for a given UI. Hassenzahl's et al [17] instrument could be used to measure hedonic quality, because it is a validated, freely available, short, and easy-to-understand verbal scale. The instrument is a seven point semantic differential scale (outstanding-second rate, standard-exclusive, impressive-nondescript, ordinary-unique, innovative-conservative, dull-exciting, interesting-boring).

5 Discussion

The main contribution of media studies to the universal access paradigm is a detailed consideration for: 1) an important class of users (the TV audience), 2) the domestic leisure time and 3) the uses and gratifications of TV and audiovisual content. Universal access has been evolving from users with disabilities and Health Telematics to children and learning in order to address the diversity of the user populations reached by contemporary computer applications. Accordingly, the universal access toolset could be further extended with the contribution of media studies (Table 3). Although media studies theory offers significant contributions and ideas, it could not form the basis for the UI evaluation methodology itself, because its research methods are not focused on improving interactive media during the development phase. In addition, media studies methods are not relevant to the design part of universal access research, although they were helpful in the conceptualization of the affective quality of an ITV UI.

Table 3. Media studies contribution to the universal access toolset

Universal Access	Traditional focus	Media studies contribution
Users	Disabled, Aged, Children	TV viewers
Application domain	Health, Learning	Entertainment
Context of use	Organization, Domestic (shopping, health)	Domestic (leisure time)

In this paper, we identified and suggested constructs that are relevant to the contemporary issues in ITV UI design and we presented respective measurement instruments that are easy to administer and compatible with the contemporary UI evaluation methods. These data collection techniques are based on self reports and have been validated either in the context of computer applications or TV studies. In this way, the affective quality of an ITV UI could be measured cost-effectively in the context of the usability engineering lifecycle [31]. Affective quality concepts could also be employed in quantitative evaluation during the adoption of ITV applications from the audience, with surveys and longitudinal studies such as the "Experience Sampling Method" [21].

The emphasis on an affective methodology for ITV applications does not entail a complete abandonment of the efficient usability paradigm. For example, an ITV news application used in the morning before leaving home for work should afford efficient information retrieval and navigation. The same application, used in the evening after returning home from a long day at work, should be more automated and encourage relaxed use (Steve Draper, personal communication). In general, the UI evaluation should be regarded to have both an affective and an efficiency dimension. In the ITV case, the leisure context of use and the need for gratifying entertainment goals might push the balance towards the affective dimension of the UI. UI developers should explicitly set the goals of each UI depending on the nature of the ITV application, and then they should employ the appropriate assortment of efficient usability and affective quality methods for evaluation.

Besides media studies, there might be alternative approaches for conceptualizing the affective dimension of an ITV UI. In this paper, we employed concepts from mass communication, in order to facilitate the universal access to ITV applications. However, ITV applications are supposed to offer more than just an improved version of the traditional TV experience. One potential benefit of ITV applications would be the creation of optimal experiences through flow [4], which requires the establishment of a match between the viewer skills and the challenge posed by the ITV application. There are also a few additional HCI paradigms that should be investigated in the context of ITV applications. For example, HCI research is gradually diversifying its focus in areas such as: 1) influencing the user through persuasion [11], which offers concepts related to the trust in advertising and commerce ITV applications, and 2) video-games and fun [7], which offers concepts related to the game-play dimension of ITV usability. Depending on the application domain (e.g., entertainment, learning, e-commerce, game-play, information) the design and evaluation of ITV applications should employ the most suitable concepts.

6 Conclusion and Further Research

The introduction of digital TV (DTV) brings the promise to enhance the TV experience and to bring the Information Society services to the majority of consumers [26]. Nevertheless, the adoption rate of DTV has been slow and the viewers' attitude indifferent towards the interactive services [43]. The success of a new technology obviously depends on many factors [29]; here we examined the role of the UI in ITV applications. Instead of adopting the traditional usability concepts and techniques, we examined the theory and the methods from media studies. In this way, the proposed

affective UI evaluation methodology takes into account the unique characteristics of the TV medium, TV audience, and context of use. Moreover, we provided a generic set of appropriate constructs and instruments for evaluating the ITV applications. Thus, the proposed affective evaluation methodology for ITV ensures that the ITV application is not only accessible and usable, but can also successfully compete with the established TV experience.

In further research, ITV applications should be assessed against the established benchmarks of the TV experience that have been described in this paper. In this way, new applications will support familiarity and acceptability for all TV viewers. In addition, the emotional responses at the visceral, behavioural, and reflective levels could be transferred to other affective UI evaluation studies, beyond ITV applications. Besides evaluation methods, further theoretical research should consider the development of UI design methods and techniques based on the affective conceptualization of an ITV UI. The ultimate objective of this work is to achieve a shift in the mentality of ITV application design and evaluation, so that the ITV user is regarded firstly as a viewer and then as a user.

Acknowledgements

7   References

[1] Berglund A, Johansson P (2004) Using speech and dialogue for interactive tv navigation. *Universal Access in the Information Society* 3(3-4):224–238

[2] Bradley M, Lang P (1994) Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therepy and Experimental Psychiatry*, 25(1):49–59

[3] Chorianopoulos K, Spinellis D (2004) Affective usability evaluation for an interactive music television channel. *Comput. Entertain.*, 2(3):14

[4] Csikszentmihalyi M (1991) *Flow: The Psychology of Optimal Experience*. Perennial New York

[5] Davis F (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340

[6] Desmet PM (2003) Measuring emotions: Development and application of an instrument to measure emotional responses to products. In Blythe, M., Monk, A., Overbeeke, K., and Wright, P., editors, *Funology: from usability to enjoyment*. Kluwer

[7] Draper SW (1999) Analysing fun as a candidate software requirement. *Personal and Ubiquitous Computing*, 3(3):117-122

[8] Drucker SM, Glatzer A, Mar SD, Wong C (2002) Smartskip: consumer level browsing and skipping of digital video content. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp 219–226

[9] Eronen L (2001) Combining quantitative and qualitative data in user research on digital television. In *Proceedings of PC HCI 2001*. Typorama, Athens

[10] Eronen L, Vuorimaa P (2000) User interfaces for digital television: a navigator case study. In *Proceedings of the Working Conference on Advanced Visual Interfaces.* pp 276–279

[11] Fogg B (2002) *Persuasive technologies: Using computer power to change attitudes and behaviors.* Morgan Kaufmann, San Francisco

[12] Freeman J, Lessiter J (2003) Using attitude based segmentation to better understand viewers' usability issues with digital and interactive tv. In *Proceedings of the 1st European Conference on Interactive Television: from Viewers to Actors?*, pp 19–27

[13] Frokjer E, Hertzu, M, Hornb K (2000) Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *Proceedings of the SIGCHI conference on Human factors in computing systems.* pp 345–352

[14] Gill J, Perera S (2003) Accessible universal design of interactive digital television. In *Proceedings of the 1st European Conference on Interactive Television: from Viewers to Actors?* pp 83–89

[15] Goleman D (1995) *Emotional Intelligence.* Bantam New York

[16] Hassenzahl M (2005) The quality of interactive products: Hedonic needs, emotions and experience. In Ghaoui C (ed). *Encyclopedia of Human Computer Interaction.* Idea Group London

[17] Hassenzahl M, Beu A, Burmester M (2001) Engineering joy. *IEEE Software*, 18(1):70–76

[18] Hassenzahl M, Platz A, Burmester M, Lehner K (2000) Hedonic and ergonomic quality aspects determine a software's appeal. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* pp 201–208

[19] Holbroock MB, Hirschman EC (1982) The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of Consumer Research* 9:132–140

[20] Knobloch S, Zillmann D (2002) Mood management via the digital jukebox. *Journal of Communication* 52(2):351–366

[21] Kubey R, Csikszentmihalyi M (1990) *Television and the Quality of Life: How Viewing Shapes Everyday Experiences.* Lawrence Erlbaum New Jersey

[22] Lavie T, Tractinsky N (2004) Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Hum.-Comput. Stud.*, 60(3):269–298

[23] Lee B, Lee RS (1995) How and why people watch tv: Implications for the future of interactive television. *Journal of Advertising Research* 35(6):9–18

[24] Livaditi J, Vassilopoulou K, Lougos C, Chorianopoulos K (2003) Needs and gratifications for interactive TV applications: Implications for designers. In *Proceedings of the HICSS 2003 conferece* p 100b

[25] Macdonald N (2004) Can HCI shape the future of mass communications? *interactions* 11(2):44–47

[26] Maguire M (2002) Applying evaluation methods to future digital tv services. In Green, W. and Jordan, P., editors, *Pleasure with products beyond usability,* Taylor and Francis London, pp 353–366

[27] Malone TW (1982) Heuristics for designing enjoyable user interfaces: Lessons from computer games. In *Proceedings of the 1982 conference on Human factors in computing systems*, pp 63–68

[28] Monk A (2000) User-centred design: the home use challenge. In Sloane A, vanRijn F (eds) *Home informatics and telematics: information technology and society.* Kluwer Boston, pp 181–190

[29] Moore GA (1991) Crossing the Chasm. HarperColins New York

[30] Murry JP, Lastovicka JL, Singh SN (1992) Feeling and liking responses to television programs: An examination of two explanations for media-context effects. *Journal of Consumer Research* 18(3):441–451

[31] Nielsen J (1994) *Usability Engineering.* Morgan Kaufmann San Francisco

[32] Nielsen J, Levy J (1994) Measuring usability: preference vs. performance. *Commun. ACM* 37(4):66–75

[33] Norman DA (2004) *Emotional Design: why we love (or hate) everyday things.* Basic Books New York

[34] O'Brien J, Rodden T, Rouncefield M, Hughes J (1999) At home with the technology: an ethnographic study of a set-top-box trial. *ACM Transactions on Computer-Human Interaction (TOCHI)* 6(3):282–308

[35] Park CW, Young SM (1986) Consumer responses to television commercials: The impact of involvement and background music on brand attitude formation. *Journal of Marketing Research* 23(2):11–24.

[36] Perse EM (1990) Media involvement and local news effects. *Journal of Broadcasting and Electronic Media* 34(1):17–36

[37] Petersen MG, Madsen KH, Kjaer A (2002) The usability of everyday technology: emerging and fading opportunities. *ACM Transactions on Computer-Human Interaction (TOCHI)* 9(2):74–105

[38] Reeves B, Naas C (1996) *The media equation: How people treat computers, television and new media like real people and places,* Cambridge University Press/CLSI New York

[39] Rubin A (1983) Television uses and gratifications: The interaction of viewing patterns and motivations. *Journal of Broadcasting* 27(1):37–51

[40] Russell JA, Mehrabian A (1977) Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11(3):273–294

[41] Stephanidis C, Akoumianakis D (2001) Universal design: towards universal access in the information society. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pp 499–500

[42] Thayer RE (1986) Activation-Deactivation Adjective Check List (AD ACL): Current overview and structural analysis. *Psychological Reports*, 58, 607–614

[43] Theodoropoulou V (2002) The rise or the fall of interactivity? digital television and the "first generation" of the digital audience in the UK. In *Proceedings of the RIPE@2002 Conference — Broadcasting and Convergence: Articulating a New Remit,* Finland

[44] Tractinsky N (1997) Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp 115–122

[45] Tractinsky N, Katz A, Ikar D (2000) What is beautiful is usable. *Interacting with Computers* 13:127–145

[46] Vorderer P (2000) Interactive entertainment and beyond. In Zillmann D, Vorderer P (eds) *Media entertainment: The psychology of its appeal*, Lawrence Erlbaum Associates Mahwah, New Jersey London, pp 21–36.

[47] Vorderer P (2001) It's all entertainment—sure. but what exactly is entertainment? communication research, media psychology, and the explanation of entertainment experiences. *Poetics* 29:247–261

[48] Vorderer P, Knobloch S, Schramm H (2001) Does entertainment suffer from interactivity? the impact of watching an interactive tv movie on viewers' experience of entertainment. *Media Psychology* 3(4):343–363

[49] Wittenburg K, Forlines C, Lanning T, Esenther A, Harada S, Miyachi T (2003) Rapid serial visual presentation techniques for consumer digital video devices. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pp 115–124

[50] Zaichkowsky JL (1985) Measuring the involvement construct. *Journal of Consumer Research* 12:341–352

[51] Zhang P, Li N (2005) The importance of affective quality. *Comm. of the ACM* 48(9):105-108

[52] Zillmann D (2000) The coming of media entertainment. In Zillmann, D. and Vorderer, P., editors, *Media entertainment: The psychology of its appeal*, Lawrence Erlbaum Associates Mahwah, New Jersey London pp 1–20

[53] Zillmann D, Bryant J (1985) *Selective exposure to communication*. Lawrence Erlbaum Associates Hillsdale

Appendix – Example Application

In this section, we provide a brief overview of an example application for one of the proposed user interface evaluation concepts [3].

The objective of the study was to evaluate user preferences for an ITV application that offers clip skipping for music video television and an animated character for presenting information. We chose to use the affective quality instrument of Hassenzahl et al [17], because it is validated, freely available, short, and features an easy-to-understand verbal scale[*]. Furthermore, a fulfilling television experience depends on the subjective evaluations of the entertaining value of the content, a characteristic that is partially captured by the construct of hedonic quality. The experiment was designed to address two of the main issues that have been identified in ITV user interface design: (a) local storage navigation through simple video clip skipping and (b) presentation of related information through alternative presentation styles. We formulated the objectives of the study as research hypotheses.

- Hypothesis 1: Hedonic quality will be greater for a clip-skipping music TV channel compared with a fixed one.

- Hypothesis 2: Hedonic quality will be greater for an animated character compared with a transparent information box for the presentation of related information.

---

[*] We used a seven-point semantic differential scale and reversed the polarity of every other pair: outstanding-second rate, standard-exclusive, impressive-nondescript, ordinary-unique, innovative-conservative, dull-exciting, interesting-boring. Scores were summed and then scaled from 0 to 10.

Each participant received two experimental treatments (within groups) of the user interface for interactive music video television: 1) The animated character and 2) the transparent box, while both of setups offered video clip skipping with ad insertion. After the end of each session, participants evaluated separately the hedonic quality of (a) traditional music video television (all participants were selected to be frequent viewers of music TV), (b) music video television with clip skipping, (c) information presentation with the transparent box and (d) information presentation with the animated character. We ran tests with 21 users (recruited from the postgraduate and undergraduate departments of our university). Ages were between 22 and 35 (13 men and 8 women). Users were assigned with a random order to each treatment and the order of the music video clips was also randomized for each session. The video clip related information and the remote control were the same for all sessions.

The study was performed in a relaxed setting, using a traditional TV set and a remote control. The testing session contained 16 video-clips and advertising breaks with three ads every 4 songs (approximately every 15 minutes), just like a commercial music video television channel. The study followed the selective-exposure paradigm. Users were free to choose the music video clip they preferred to watch, like they would do if the experiment was not running. In order to ensure selective-exposure, the users were allowed a maximum of 1/3 of watching time, out of the total session duration, that is a maximum of approximately 20 minutes out of the 1-hour program duration. Users could press the power-off button on the remote to end the testing session and they were told to watch as much as they liked, between 10 and 20 minutes.

We found (Table 4) that the hedonic quality score (scale is from 0 to 10; scores less/more than 5 represent negative/positive attitude) for the traditional setup is close to neutral (average 5.1/10). This finding can be explained by the fact that music video television is a pervasive experience and feels familiar to consumers, irrespective of its delivery format. In contrast, video clip-skipping (average 7.5/10) allowed experimental subjects to watch favourite music video clips and despite the dynamic insertion of ads the hedonic quality score was significantly higher (two tailed t–test, p=0.002, n=21). Therefore, we argue that simple video clip skipping, similar to the track-skipping facility available in audio CD players, enhances the perceived television entertainment value, when compared with same fixed TV channel.

Table 4. Mean hedonic quality scores for the clip-skipping music video television are significantly higher

Hedonic Quality (p=0.002, n=21)	Average	Std Dev
Music TV (traditional)	5.1	2.1
Clip-skip	7.5	1.6

Consistent with the selective exposure theory, users actively sought for the video clips and songs they preferred. This kind of interactive behaviour may be due to the experimental setting and may not have external validity; users may have been more engaged than normal because the application was novel to them and because they were specifically asked to use the new system. They reported that they used the skip functionality mainly to skip a music video that they disliked and to a lesser extent to get to a favourite one. Either way, the clip-skipping feature was liked, despite the ad insertion, and provided a relaxed way to control the interactive music TV application.

We also found that the hedonic quality (scale is from 0 to 10; scores less/more than 5 represent negative/positive attitude) for a music video television channel is significantly higher (two tailed t–test, p=0.0002, n=21) when using an animated character for presenting dynamic video overlays (average 7.0/10) compared with the traditional transparent information box (average 4.4/10). Again, the experimental subjects were neutral toward the traditional information box, since it is a widely used and familiar presentation style for information related to music video clips (Table 5). Therefore, we

argue that the animated character could be used to enhance the consumers' entertainment experience with television.

Table 5. Mean hedonic quality scores for the animated character compared to the traditional overlay box

Hedonic Quality (p=0.0002, n=21)	Average	Std Dev
Animated Char.	7.0	1.5
Box (traditional)	4.4	2.0