공학박사 학위논문

A mixed-method approach to measuring interactive TV UX: Study of conversational agent equipped TV

복합적 방법을 활용한 TV 사용자 경험 측정

-대화형 에이전트 TV를 중심으로-

2020년 2월

서울대학교 대학원

산업공학과

이 승 연

ABSTRACT

A mixed-method approach to measuring interactive TV UX: Study of conversational agent equipped TV

Seungyeon Lee
Department of Industrial Engineering
The Graduate School
Seoul National University

This is a study on the interactive TV user experience (UX). In the field of human-computer interaction (HCI), system performance or usability has been evaluated as a major factor affecting user satisfaction. This approach was very effective when the system had been used in specific area, or when the range of user-computer interaction was limited. However, with the development of technology, we can now freely choose and use these technologies. This indicates that users expect more than usability for technology. Some of the technologies survive for a long time or contribute to enriching our lives, while others disappear quickly. Because of the uncertainty and complexity of new technology, industry is now specially focusing on users' motivation and experiences than on the technology itself (Shin, 2009; Thong et al., 2006). The user experience refers to a comprehensive experience that includes emotional dimensions such as fun and pleasure as well as performance and usability.

TV is one of the most accessible entertainment products that provide pleasure at low cost in our daily lives. As the TV platform has been combined with the Internet and various functions can be accommodated, TV is evolving into the interactive product. In practice, however, various functions of interactive TVs are not actively used and utilized. For the reason why they are not used, unnecessity and inconvenience are cited. The purposes of this study are to (1) identify main attributes affecting the interactive TV UX, (2) evaluate alternative user interfaces for enriching the TV user experience, and (3) construct an interactive TV UX model framework to see the causal relationship between the user experience, user satisfaction, and intention to use technology. Before the experiments, previous studies and important claims to the UX theories were reviewed and analyzed. For the first purpose of this study, a lab study was conducted to observe and compare the user's behavioral situations in watching TV. The user behaviors were measured in two different situations - the goal-directed situation and the experiential situation. For the second purpose of this study, an empirical study was performed to observe whether user experiences differ between remote control unit (RCU) interaction and conversational agent (CA) interaction, when considering an alternative interface for interactive TV. Two methods – subjective self-reporting questionnaires and physiological measurements for user's emotional response – were used. The field test was also conducted for 20 households for 4 weeks. Participants assessed their own moods and recorded the TV viewing experiences. User satisfaction and the continuous usage intention were measured once a week to see whether the user satisfaction and the usage intention show the relationship each other. Finally, the structural equation model was constructed to validate the UX measurements scale and examine the causal relationship among the experience quality factors at each UX phases. A specific relationship between an interactive TV UX framework and how UX quality factors

affects the user satisfaction and the decision to use the technology was also confirmed.

The results from the series of studies are as follows: The experience of TV users differs depending on the context of the situation. When watching TV for a specific purpose, the quality of the experience was higher and more positive than when watching TV with no purpose. The factors affecting TV user experience are “controllable,” “innovative,” “impressive” and “original”. In the user interface comparison experiment, quality factors affecting the user experience were “manageable,” “predictable,” “stylish,” and “innovative.” Physiological measures were conducted to identify real-time emotional responses that occurred during the interaction. The result was also confirmed that the user's cognitive effort and stress appeared differently according to the interfaces, RCU and CA. As a result of skin conductance (SC) measurement, CA interaction showed mainly high to low in the level of stress, while the result of RCU showed low to high or no change in stress. This implies that the user is likely to feel a cognitive load at the beginning of interacting with CA. Conversely, controlling with RCU method may not require a large cognitive burden, but stress is increased in navigating menus and clicking buttons repeatedly. This shows that the roles of CA and remote controller are different, and therefore, CA is suitable for complementary role rather than alternative means of RCU in interacting with TV. In this regard, it is necessary to develop the TV interface, which is optimized for the usage of RCU only, to the interface considering both the CA and the RCU interaction to enhance the TV UX. Field tests showed a steady increase in user satisfaction for four weeks compared to the pre-evaluated expectation. The users’ intention to use decreased for three weeks and then increased again at the fourth week. As a result of reviewing the comments of user’s diaries, it was found that users’ complaints occurred in communication errors were

very high at the beginning of the experiment. It was also seen that users have learned and accepted CA through the repeated daily use. Finally, it was constructed the interactive TV UX model framework through the structural equation modeling and confirmed the quality of interactive TV user experience. UX quality factors affecting the overall user experience are marked with “manageable,” “predictable,” “stylish,” “inventive,” “innovative,” and so on. The common factors derived from each experiment are “manageable,” “innovative,” “captivating,” and “novel.” As a result of confirmatory factor analysis, it was validated that the positive use experience quality has the causal relationship with the use satisfaction, and the use satisfaction also correlates with the continuous usage intention. The hedonic quality of user experience more affected user satisfaction and intention to use than to the pragmatic quality. Although CA is technically unfinished, it is more appealing to users in terms of emotional quality, predicting that voice-based communication may become a new paradigm of human-computer interaction. This study confirms that not only the user's experience and satisfaction can be the meaningful measures of the continuous usage intention new technology, but also a positive user experience is essential for the success of the technology. It is also seen that a mixed-method approach allows rich implications by examining the interactive TV user experience from various perspectives. It is hoped that this research can be used as a reference for providing a more positive user experience in the development of TVs and other interactive systems using CA technology.

Keywords: interactive TV, user experience, everyday products, conversational agent, hedonic quality, ergonomic quality

Student number: 2011-31008

CONTENTS

List of Table

List of Figure

Chapter 1
Introduction

1.1 Background and motivation

Currently, new technologies have characteristics that are complex and difficult to define; for example, they have multiple attributes simultaneously or converge different types of products and services. Because of the uncertainty and complexity of new technology, industry is now specifically focusing on users' motivation and experiences than on the technology itself (Thong et al., 2006; Shin, 2009). We live in a user-centric era where people freely choose and use products/services in every part of our daily lives. Many technologies are embedded in consumer products to enrich our lives. Technologies that people don't use disappear quickly. The proliferation of new technologies occurs when users decide that the uncertain benefits exceed the uncertain costs of accepting the technology (Venkatesh et al. 2000). Therefore, the viability of new technologies and evolved products both depends on the choice of the users and the continuous usage intention. The continuous usage intention plays an important role in the long-term viability (Van Schaik and Ling, 2008) and ultimate success of technology (Thong et al., 2006). Exploring whether users have an intention to use is necessary to predict the future of new technology.

The key factor related to intention to use is user satisfaction (Deng et al., 2010; Lee and Choi, 2017). Chung and Kwon (2009) noted that mobile experiences and technical support were factors that contributed to the intention to use in their mobile

banking study. Deng et al. (2010) studied intention to use in IT and argued that user satisfaction exerts a positive effect on intention to use. In a study on mobile users, Kim et al. (2013) argued that user-engaged motivation affects user satisfaction and engagement intention. User satisfaction have been the ultimate goal in the human-computer interaction (HCI) field. For a long time, researches in HCI field have been paying attention to cognitive aspects of human factors and usability for user satisfaction.

Such a perspective has been regarded as an appropriate assessment in the era when it places importance in the usability aspects like productivity and the machine performance. For a long time, the fulfillment of these requirements was considered sufficient for the user's needs. For example, robustness, ease of use and safety have been considered as advantages of consumer electronics such as TV and refrigerator. The interaction and experience between the user and the system were not the main issue. With advances in technology, legacy products were evolving to the interactive systems and the scope of user experience covers all systems that can be interacted with, from games to everyday products. People began to pay intention to other qualities beyond the utilitarian qualities. Interactive product needs a comprehensive user experience study that takes into account the diverse interactions between the users and the technology and it is expected to link to the better user satisfaction.

UX is defined as the every aspect of the user's interaction with a product (ISO 9241-210(2010). UX studies have been conducted in various ways in many areas, and there are many opinions regarding their scope and effects (Brajnik & Giachin, 2014; Law & Sun, 2012; Mahlke, 2008). Inspite of this, it is a common characteristic that the emotional aspects are emphasized as important factors of UX (Gaver & Martin, 2000; Hassenzahl, 2001; Hassenzahl & Tractinsky, 2006; Jordan, 2000; Lenderman, 2006; Mahlke & Thüring, 2007). Norman (2004) claimed that the effect of emotional impact on the UX is larger than that of the usability. Chitturi, Raghunathan and Mahajan (2008) viewed that people focus choices on the design and exterior appearance instead of functions and performance that are not clearly visible in the

outside. Tractinsky, Shoval-Katz and Ikar (2000) asserted that the aesthetics factor is evaluated by being closely combined with the usability, and while emphasizing the importance of emotional factors. In 2002, based on a result that the viewers liked more the user interface (UI) that incurred more time and errors in a video skipping UI assessment, Drucker et al. (2002) stated that the subjective satisfaction of users was different from the efficiency measurement standard. Ardito et al. (2008) stated that experiences are influenced "by the user's psychological state and the context" as well as the characteristics of the interactive system. According to UX related studies, quality attributes of UX are generally known as two sub-dimensions, pragmatic quality (PQ) and hedonic quality (HQ) (Hassenzahl, 2003; Mahlke, 2006). The overall user experience is based on the combination of these two quality dimensions. Pragmatic quality includes usability oriented attributes such as predictable, utility, and manageable. Hedonic quality indicates emotional attributes such as fun and excitement. The extent of user experience research available covers all systems that surround us, from games that require high concentration to everyday products.

1.2 The purpose of this study

The main purpose of this study is to explore what are the main UX attributes affect the user satisfaction in TV interaction. Television (TV) has been loved for a long time as an entertainment medium for relaxation and fun. As TV has become connected to the Internet, and the TV platform has developed to accommodate a variety of applications, it is increasingly becoming an active and interactive medium and the experience of TV viewers is also changing. This diverse function and active usage strengthened by industrial developments obviously differentiates interactive TV (or smart TV) from the previous non-interactive TVs (hereinafter "legacy TV"). The industries expected TV viewers are more engaged in TV and actively use new features of TV. They also predicted that TV would serve as the center of home

appliances or the main platform of consuming products. In reality, however, the TV viewer's behavior and satisfaction to the interactive TV are not much different from those to the legacy TV. Not many studies convince that interactive TV enhances the TV viewer's satisfaction. Instead, there are contrary studies that TV viewers rarely use the interactive TV's new features because they are unnecessary or difficult to use (Darnell 2007; Lee and Lee 1995). The reason is that it was focused on new business opportunities (Bellman et al. 2017; Lekakos et al. 2001) or its technological aspects (Morris and Smith-Chaigneau 2012), rather than investigating TV viewer's entertainment needs (Chorianopoulos and Spinellis 2004; Livaditi et al. 2003) or enhancing the experience of interactive TV viewers.

Also, there is not enough comprehensive research on TV user experience from the user's point of view. Chorianopoulos and Spinellis (2004) argue that the evaluation of traditional user interfaces may not be suitable for evaluating interactive applications so that TV viewers' emotional experience needs to be measured differently. As mentioned earlier, new technologies and systems are spreading and evolving when users are actively and naturally using them. Exploring TV UX from the user's point of view may be the first step to provide the expanded user experience and enable TV more usable and enjoyable. User experience is not always the same, but differs depending on system characters, user personal characteristics, and the context which interactions occur between user and system (Hassenzahl and Tractinksy, 2006; Mahlke and Turring, 2006). Marketing and Consumer Behaviors, consumer related studies, suggest do-goal and be-goal as representative contextual situations. In do-goal, PQ have a greater impact on the user experience. In be-goal, users are more influenced by emotional quality. Experiment comparing interactive TV user experience in different behavioral situations – do-goad and be-goal - is conducted. It will be examined whether interactive TV users are experienced differently in two behavioral situations. In addition, experiments will reveal in which situations the users are appealed more positively.

The second purpose of this study is to investigate which user interface may provide

a more extended user experience in the TV interaction. When iTV first appeared, it showed limited performance due to constrained hardware and bandwidth (Jensen, 2008). These hardware constraints have been improving rapidly due to evolution of Information and Communications Technology (ICT). As TV evolves into interactive media, there are also a number of technological changes surrounding it. However, despite technological changes, most TV interactions are handled by the remote control, which is not adequate for interactive technologies. RCU is the most optimized user interface for the lean back environment. While interactive environments may enable more active interactions, the RCU is unable to accommodate the variety of activities that occur in these interactive environments. On the other hand, it is true there is no alternative that enhance the interaction beyond the RCU. As a result, many of the features and services in TV are not used and experienced. Interface optimization may the key to enrich the user experience of interactive TV.

Recently, conversational agent (CA), one of the new technologies combined with TV, has been discussed as one of the alternatives for controlling interactive TV features. Based on the voice recognition technology, CAs, such as the recently developed Apple Siri, Amazon Alexa, and Google Assistant, provide information including weather data and assist tasks in our daily lives such as playing music. Furthermore, by combining it with a device such as TV, the CA performs the role of an assistant that helps the user. It is obvious that without changing the TV viewing context, the CA assists a user who feels uncomfortable with the complex TV control function (e.g., text input and content search). CAs appeared in the late 2000s. Because of low voice recognition rate and insufficient database, the CA technology was advancing at a lower rate. In recent years, the CA functions have been improving at a faster pace through technological advancements such as machine learning and cloud-based big data accumulation. It is necessary to explore the possibilities for voice recognition technology that may be an alternative interface for interactive TV in the near future. Two empirical studies are performed to explore two interfaces (RCU and

CA). First, lab study is conducted to evaluate the user experience while users are interacting with TV, as well as after interaction. At this time, in order to objectively verify the results of the UX measurements, biological measurements and post-interaction evaluation are performed in parallel, and the correlation between the two research results is measured and analyzed. In addition, an appropriate method for UX measurement is discussed. Usually the measurement method used in UX studies is a subjective evaluation method, in which the user evaluates his or her own feeling or mood after interacting with the system.

While it is valid for the user to directly assess his or her own condition, it is less reliable, because it is a subjective self-assessment. Second, a real-life-condition experiment was conducted. An empirical study of cumulative UX using an interactive TV was conducted. Households with interactive TVs were asked to record and evaluate their experiences while watching TV every day for four weeks. In the field of human-computer interaction (HCI), user experience indicate the experience arising from direct interaction with the system. The term of 'experience' refers to a moment as well as a series of process in which users evaluate and respond to the system cognitively and emotionally. User experience occur all at once. It can also change in time and is accumulated over and over again. Positive emotions are known to increase over time (Norman, 2004; Chitturi et al., 2008). Most UX related researches are, however, often one-time or short-term studies. Especially for new technologies that have not yet been used or are not familiar to users, or for evaluating the user experience as a measure of predicting continuous usage intention, it is important to consider it may take time for the users to be accustomed to the technology.

The third purpose of this study is to develop the interactive TV UX research model and confirm whether the perceived user experience affect the acceptance of new technologies. Technology Acceptance model (TAM) (Davis, 1987) which are often used for predicting intention to use technology, are known to be suitable for usability-

oriented systems or early stages of technology acceptance. Existing UX studies generally focus on the UX process or overall framework for the attributes of UX quality dimension, rather than investigating the influence of user experience to the user's behavioral intention to use or relationship among them. Many studies describe the user satisfaction and the behavioral intention as the overall evaluation together. These are rather abstract and closer to phenomenological studies.

In this study, Structural equation modeling is adopted to identify whether the user experience has a causal relationship between UX qualities, user satisfaction, and continuous usage intention. By predicting this, researching the user's point of view may not be a mere phenomenal research, but be a practical effort supporting the specific design direction or business strategy. In the 21st century, technologies are freely applied to the non-utilitarian products as well as utilitarian products. It is necessary to perform a holistic investigation of user experience, user satisfaction, and the prediction of the technology's prolonged use. This may contribute to making the product interactions that are being made rapidly become meaningful changes that extend and positively induce the user's experience.

1.3 Dissertation structure

This paper is organized as follows. Chapter 1 is the overview of this dissertation about interactive TV UX. Chapter 2 is about user experience related theories, previous studies and disciplines about UX. Chapter 3 defines the variable features and core features of interactive TV and reviews the related research on CA features. Chapter 4 compares usage experiences that vary with context. In Chapter 5, we present an experimental study on how UX differs between the CA and RCU

interfaces. In this case, both physiological measures and self-reporting questionnaires are used to confirm both the user emotion during the interaction and post-use experiences, and the correlation between the two results. In particular, we examine how the RCU experience and the CA experience affect the quality of UX and the intention of continued-use. Chapter 6 discusses the measurement of interactive TV UX in everyday life. Data were recorded over four weeks by 20 households using interactive TVs to analyze how TV experiences change over time. In Chapter 7, structural equations are used to evaluate the causal relationships that affect the evaluation of UX, using the results of the previous studies. In particular, we examine the relationship between expectation, experienced user satisfaction, and intention of continued use. We also analyze how the experience quality affects the evaluation of the experience, the validity of the measurement variables, and the causal relationship between important variables. Finally, Chapter 8 presents the major discussions, contributions, and the scope for further studies.

The objectives of this study are as follows. First, we define the interactive TV UX and discover key experience quality and measurement metrics. Second, we identify the most appropriate measurement method based on the experience-evaluation goal. Currently, subjective user experience evaluations, such as self-reporting (wherein users measure their experience themselves), is used; however, subjective measurement methods may have different evaluation criteria for each user, and thus, it is necessary to secure objective data as well. We will discuss alternative methods to complement subjective measurement methods and the benefit of using multiple methods. Third, we evaluate if the CA, one of the new technologies, can be established as an interactive TV interface based on UX. Finally, a TV UX framework is proposed by verifying the adequacy of the quality variables used for measuring UX and the causal relationship between major variables for each experience level.

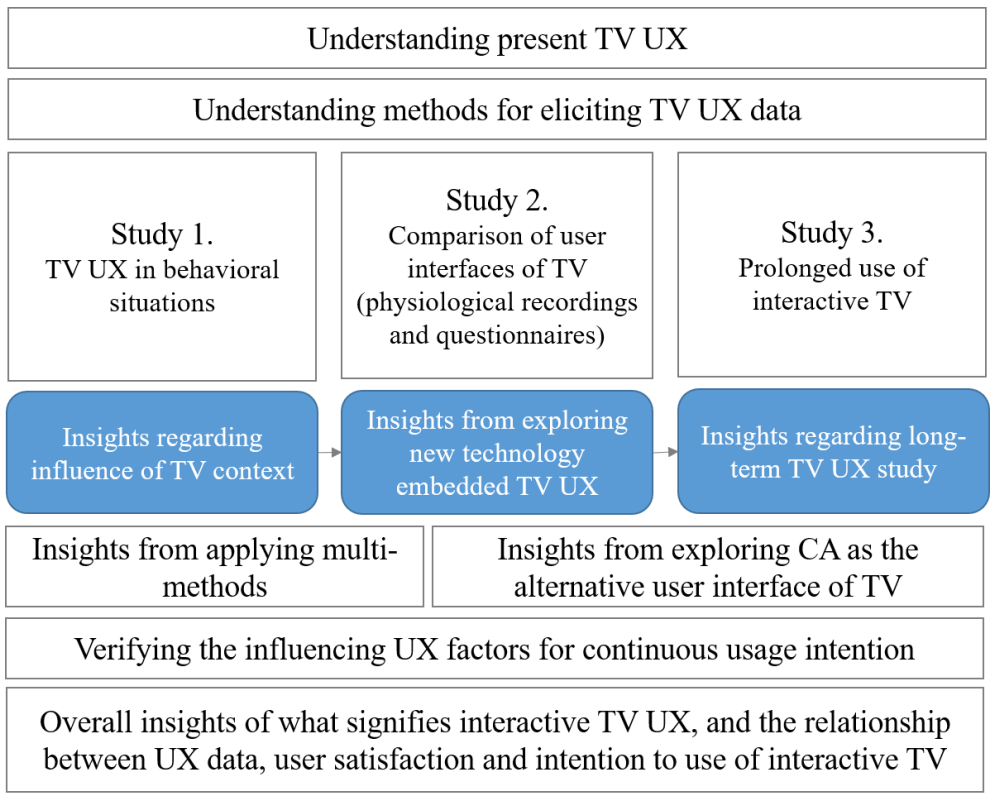


Figure 1.3.1. The flowchart of this dissertation

Chapter 2
Literature Reviews

Beyond usability in HCI

Understanding the needs of users is the first, but difficult task to step in product development (Chapanis, 1995). For a long time, Human-Computer Interaction (HCI) has mainly focused to cognitive aspects of human factors and usability analysis for user satisfaction. Such a perspective has been regarded as an appropriate assessment method in the case of a system that places importance on the productivity and performance. The HCI goal, defined in ISO 9241-210, is to provide an "optimal user experience through availability, usability and satisfaction". From the perspective of the computer, as a tool for achieving the user's desired purpose easily and quickly, availability and usability are very important factors for user satisfaction.

There are other claims that non-utilitarian factors, such as fun, enjoyment, or aesthetics, are the main factors that affect the user experience (Hartmann et al. 2007; Norman 2004; Tractinsky et al. 2000; Tuch et al. 2012). Tractinsky et al. (2000) argued that the more beautiful the design, the more useful the result. Norman (2004), who emphasized the importance of emotional factors, argued that non-utilitarian aspects play a mediating role in goal-directed activity rather than usability, because the emotional impact through the aesthetic factor has a greater effect on the user experience. Dillon (2002) insisted that aesthetics work as a "driver," enhancing

usability. Wang et al. (2010) insisted that web aesthetics plays a substantioal role in activating online shoppers' information processing.

Usability and emotion are associated with the cognition and affect attributes, respectively, and these two attributes are experienced in instrumental and non-instrumental aspects, respectively, through user-system interaction. The instrumental attribute is sometimes referred to as pragmatic (Hassenzahl, 2003), utilitarian (Batra & Ahtola, 1990), or functional (Kempf, 1999), depending on the scholars. Instrumental aspects refer to the instrumental perspective from which the human-system interaction seeks to achieve its purpose from the traditional perspective. The success or failure of the task is the most direct and clear criterion for the user's experience satisfaction, and the task success rate, a short execution time, and a low error frequency indicate high user satisfaction. All of these results can be objectively estimated through quantitative methods such as usability evaluation. Non-instrumental aspects refer to the emotional, hedonic (Hassenzahl, 2003), affective quality of a product or non-task related attributes of a product. Unlike instrumental aspects, non-instrumental or emotional aspects refer to the user's subjective feelings, moods that mostly use self-report questionnaires or diary methods.

Utilitarian and non-utilitarian aspects are also linked to fundamental 'human needs.' Human needs are changeable and subjective. Fun, excitement, appeal (Gaver and Martin, 2000), challenge, curiosity, emotional connection (Malone, 1981), surprise, diversion are products' attributes meet fundamental human needs. Jordan (2000) also referred a hierarchical organization of user needs and emphasized the importance of pleasurability for enhancing the user – system interaction as well as functionality and usability. Helander and Khalid (2006) describe the relationship between affects and cognition by borrowing Maslow's hierarchy of needs (1968). According to Maslow

(1968), humans first have deficiency needs such as safety, functionality, and usability, but once these basic needs are met, they are extended to more personal and subjective areas such as pleasurability, and individuation. Hancock et al. (2005) developed this to categorize low needs, such as safety and functionality, into ergonomics and high needs, such as surprise and fun, into the hedonomics quality (See Figure 1.3.2). In other words, user needs have multiple stages, from low performance or usability needs to overall user needs, and UX is close to a comprehensive human need that encompasses them all. Kano proposed two orthogonal dimensions, the necessary condition and the sufficient condition (Sauerwein et al., 1996). According to his research, the user is not impressed with the necessary condition, but is disappointed if it is not met.

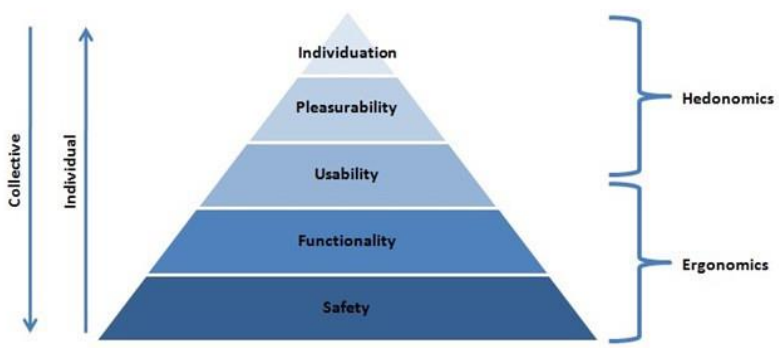


Figure 1.3.2 Hierarchy of ergonomics and hedonomics (Hancock et al., 2005, reprinted).

On the other hand, regarding the sufficient condition, users are less disappointed if it is not met, and they are much more impressed, when it is. Incorporating the Kano model into UX, usability is closer to the requirement, and user satisfaction is closer to affective user satisfaction. Zhang and von Dran (2000) see 'usability' as a kind of

'hygiene factor' (Hertzberg, 1966) and argue that when it is not met, it causes negative emotions, but increased satisfaction regarding it does not enhance the positive experience.

Usability aspects and emotional aspects have independent and opposing attributes, but at the same time the two qualities influence each other and are very closely linked. Helander and Khalid (2006) asserted that symbolic and subjective personal interpretations of the stimuli are created by cross-coupling between affect and cognition.

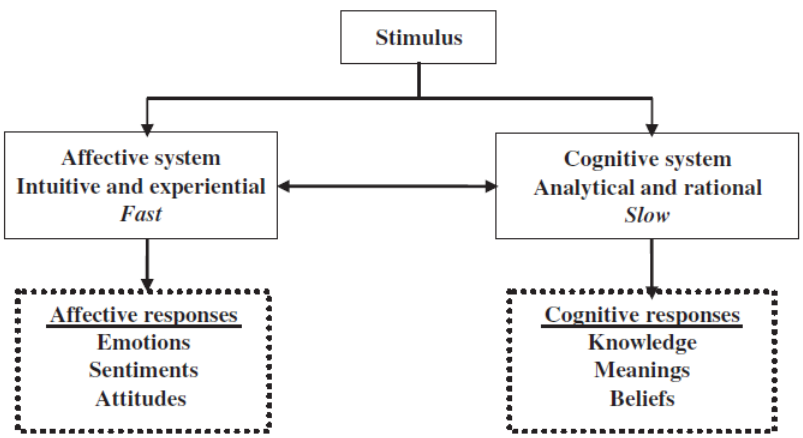


Figure 1.3.3. Helander & Khalid's affective and cognitive responses (2006, reprinted)

Kurosu and Kashimura (1995) argued that there was a strong correlation between apparent (perceived) usability and interface beauty through 26 different interface experiments on ATM devices to examine the correlation between visual aesthetics and usability. Tractinsky et al. reproduced this experiment in 2000 again. They asserted that the aesthetics factor is evaluated by being closely combined with the usability, and while emphasizing the importance of emotional factors. Zhang and Li

(2004) found the perceived affective quality of a course management system to be an antecedent of its perceived usability, usefulness and the intention to use. Norman (2004) also claimed that the effect of emotional impact on the UX is larger than that of the usability. Chitturi, Raghunathan and Mahajan (2008) viewed that people focus choices on the design and exterior appearance instead of functions and performance that are not clearly visible in the outside. Beauty goes beyond the instrumental (Hassenzahl and Tractinsky, 2006) and this kind of non-instrumental aspect enriches the product quality to create more complete, holistic HCI. According to this argument, the beauty of interactive products also means increased usability (Hassenzahl, 2004). It is like the results of a social psychology study where people believe good-looking people will be more successful. The claim that cognition and affects are interconnected has been mentioned in other disciplines. An example can be found in Damasio's 1994 study of patients with observed damage to the frontal cortex of the brain. Damasio compared patients who had intellectual capabilities but did not have a “marker” for affective information with the general population, and found that the patients did not make effective decisions compared with normal people. In the field of neuroscience, it was validated through functional magnetic resonance imaging (fMRI) that cognition and emotions are unified and contributed to each other (LeDoux, 1995).

2.1 The emergence of UX

As technology evolves rapidly, and new technologies freely combine to general products, a more comprehensive concept of UX was introduced in HCI field. This occurred because usability alone does not explain why people buy more expensive smartphones and prefer beautiful but time spending UIs (Drucker et al., 2002). As a

result, the argument that users' subjective responses result from fulfilling a different dimension (emotion or affect) than usability has begun to gain strength. The term UX comes from a number of opinions, but mostly, it is accepted that Norman and Draper first used the term to refer to today's user-centric design:

> *"This section of the book contains chapters that get directly at the question of the quality of the user's experience. This is of course the ultimate criterion of User Centered System Design, but most workers approach it obliquely in various ways such as exploring the implementation techniques, or applying existing cognitive approaches."* (Norman & Draper, 1986, p. 64)

ISO 9241-210 (2010) defines UX as "a person's perceptions and responses that result from the use and/or anticipated use of a product, system or service." According to the definition of UPA (2006), UX is the every aspect of the user's interaction with a product, service, or company that make up the user's perceptions of the whole. Hassenzahl and Tractinsky (2006) defined UX as "a consequence of a user's internal state, the characteristics of the designed system, and the context within occurring the interaction.

In the HCI domain, the affective design movement began to emerge from usability under the names Affective Computing (Picard, 2005; Picard and Klein, 2002) or Kansei Engineering (Nagamachi, 1995). The main purpose of Affective Computing is that allows computers to recognize, interpret and simulate human affects (Hudlicka, 2003; Cen et al., 2016). Kansei Engineering aims to convert the user's psychological feelings that were considered difficult to measure into measurable variables to have quantitative results (Nagamachi, 2002; Jindo and Hirasago, 1997; Schütte et al., 2004). These theories, however, still focus on looking at the subject from a computer point of view rather than the user's, focusing on short-term user

emotions, and solving or improving negative emotions rather than positive emotions (Cockton 2002; Picard and Klein 2002).

UX studies have been conducted in various ways in many areas, and there are many opinions regarding their scope and effects (Brajnik & Giachin, 2014; Law & Sun, 2012; Mahlke, 2008). Many researches have emphasized the importance of emotional aspects in UX. Roto (2007) argued that usability originates product attributes and UX is subjective and personal about the product. Schulze and Krömker (2010) defined UX as "the degree of positive or negative emotions that can be experienced by a specific user." Studies on everyday products like TV, which are more focused on non-utilitarian tasks, have begun to emphasize the importance of hedonic judgements for the overall user experience (Hassenzahl et al. 2000; Hassenzahl and Ullrich 2007; Mano and Oliver 1993). Bernhoupt and Pirker (2013) also asserted that aesthetics, emotion and stimulation are important UX factors on the interactive TV domain. The main features of the major UX studies asserting emotional aspects are summarized in the table below.

Table 2.1.1 Characteristics of UX studies

Papers	UX characteristics	UX Dimension	Research goal
McCarthy and Wright (2007)	Subjective, holistic, awareness, emotional, momentary	Anticipating Connecting Interpreting Reflecting Appropriating Recounting	Theoretical
Forlizzi and Battarbee (2004)	Subjective, conscious, emotional	Experience An experience Co-experience	Theoretical
Hassenzahl and Tractinsky (2006)	Subjective, human, judgement, emotion and affects	Beyond the instrumental Emotion and affect The experiential	Empirical
Desmet and Hekkert (2007)	Subjective, holistic, conscious and awareness, emotional	Emotional Meaning Aesthetics	Theoretical

Thüring & Mahlke (2007)	Subjective, holistic, appraisal, emotional, over a limited period time	Perception of instrumental & non-instrumental qualities Emotional user reaction Consequence of the UX	Theoretical
Law, Roto, Hassenzahl, Vermeeren, and Kort (2009)	Subjective, context-dependent, dynamic,	Before interaction During interaction After interaction	Questionnaires for UX statements and definitions
Hassenzahl, Diefenbach & Goritz (2010)	Subjective, holistic, conscious and awareness, emotional, continuous stream	Pragmatic Hedonic Overall evaluation	Empirical
UX White paper (Roto et al., 2011)	Individual Context-driven subjective	Anticipated Momentary Episodic Cumulative	Theoretical
Aranyi (2015)	-	Perception of instrumental & non-instrumental qualities Emotional response UX outcome	Empirical (Online News sites)

Commonality of these arguments is that they claim that holistic, subjective and positive aspects are main features of UX. Holistic aspects focus on task related aspects, whereas UX deals with non-task and overall aspects. Based on measurements, UX uses mainly subjective methods (e.g. self-report questionnaires). In contrast, usability is characterized by the emphasis on objective methods (e.g. eye-tracking). In addition, while usability emphasizes improvement in problem or errors of the products or negative emotions, UX research focuses more on positive outcomes (e.g. positive emotions).

There is another perspective of UX as a comprehensive concept that includes usability, not as opposed to emotions. Helander and Khalid (2006) cited Whirlpool's duet washer and dryer case, which are twice as expensive as other products, claiming that traditionally purpose-oriented products (e.g. washing machine) now also

emphasize entertaining. Blythe et al. (2007) subdivided the various studies surrounding UX into five aspects: theory, purpose, method, domain, and application. Theory is about the scope of the study of UX and has two characteristics, reductive and holistic. Reductive is the study of UX in a very limited scope, and holistic, on the other hand, refers to a very comprehensive approach. Purposes are distinguished according to whether the purpose is to study theory development or to obtain instant useful data for the development required by the industry. Methods include qualitative and quantitative research of UX. Domain distinguishes whether an interactive situation is purpose-oriented or not. Non-instrumental attributes, such as pleasure, are not largely considered when the user context is purpose-oriented. Conversely, emotional aspects are an important factor if the purpose of the use is simply for pleasure, or if the meaning is in the action itself, such as to spend time. In other words, the standard of enjoyment may vary depending on the purpose of use or user expectations. Application is based on whether the application to be analyzed is for personal experience or group experience.

2.1.1 The influencing factors of UX

The common factors in users – system interaction are user, system, interaction and context. (Nicolás and Aurisicchio, 2011). Regarding factors affective UX, Hassenzahl (2003) considered instrumental quality as pragmatic aspects and non-instrumental quality as hedonic aspects are main factors affecting UX and built a UX framework based on these two quality dimensions. He have developed the UX framework from various perspectives, such as combining human needs with social power, novelty and change (Hassenzahl, 2001), or separating hedonic qualities into sub-dimensions - hedonic quality stimulation, hedonic quality identification, and hedonic quality evocation (Hassenzahl, 2003). According to Hassenzahl's UX

research model (2004)'s user experience research model (Figure 2.1.1), people perceive interactive products in terms of both their pragmatic qualities (PQ) and hedonic qualities (HQ). Hedonic aspects and pragmatic qualities are sub-properties that describe a user's experience, and the user's overall experience is a combination of these sub-experiences. The pragmatic aspects mainly involve the product's utility and usability, including it being simple and predictable. The hedonic aspects include aesthetics, self-assessment, relatedness, novelty, and so on (Jordan 2000). They argues that the product's objective quality is judged through the user's perception in hedonic quality and ergonomic quality, and the process of making the overall judgment. In summary, UX factors are both pragmatic qualities such as efficiency, and hedonic qualities such as enjoyment, and user satisfaction is achieved through the combination of those user's perceived qualities.

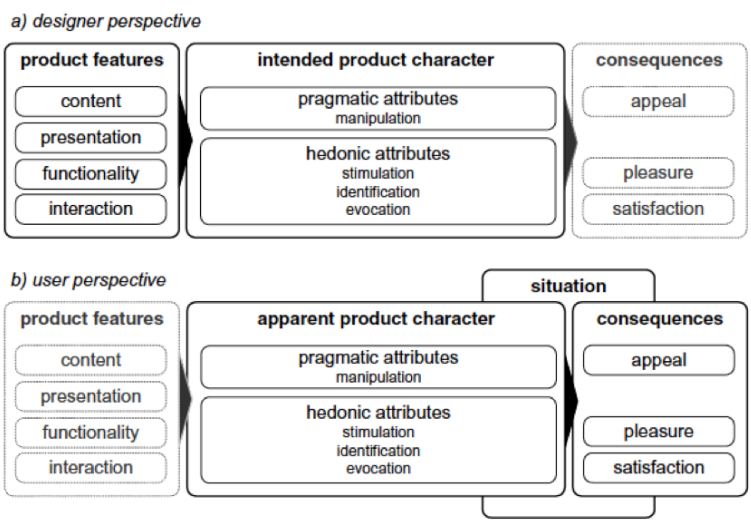


Figure 2.1.1 Hassenzahl's user experience model (2004, reprinted)

Mahlke and Thüring (2007) described a research framework for user experience and

asserted instrumental and non-instrumental quality perceptions as well as emotional user reactions as three central components of the user experience (see Figure 2.1.2).

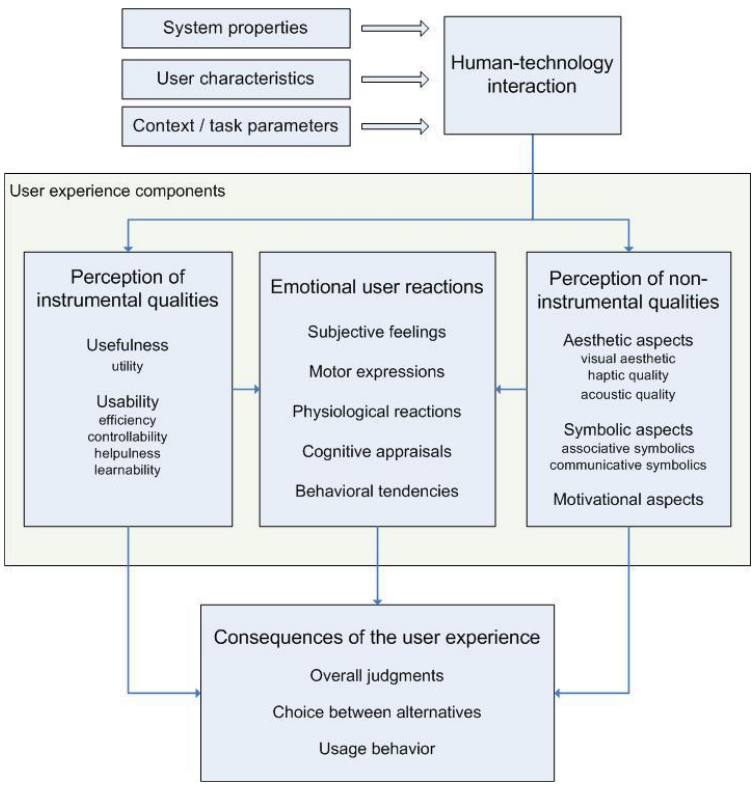


Figure 2.1.2 UX framework (Mahlke and Thüring, 2007, reprinted)

They asserted that interaction characteristics, which depend on system properties, user characteristics and context and tasks, affect these three components. The system properties are linked to user characteristics since basic human needs are key drivers of product use (Hassenzahl, 2008). According to Mahlke (2008), system properties are classified in utility, usability and visual aesthetics and human needs are from security to influence and/or relatedness. Besides system properties and user characteristics, the context of use is also one of the main variables that affect the user experience. User experience is not the same, but is influenced by variables other than

user-system interaction. For example, pragmatic quality is more important when a user interacts with a system for a specific purpose. Conversely, hedonic quality becomes a more important quality of experience when the use itself is the goal, without a target or specific task. Macleod (1994) argues that the quality of the experience depends on the application domain, context of use, and user characteristics and goals. Much research on user context has been conducted in consumer behavior research. According to Apter's (1989) reversal theory, users are affected more pragmatically when they are unconsciously or consciously exposed to a goal-oriented situation.

Conversely, if a behavior itself is the goal, the hedonic aspects become more crucial factors to the users. Many studies on consumer experiences in online environments argue that there are two types of consumption behavior: goal-directed and experiential (Novak et al. 2003; Sánchez-Franco and Roldán 2005). According to Novak et al. (2003), these two categories of behavior have distinct different attributes in important consumer behavior issues such as involvement, search and decision making. Goal-directed behaviors have the following attributes: extrinsic, directed search, and planned purchases. The user experience is affected not only by the product's own attributes but also by the user's behavior. These behaviors are mainly related to the ergonomic quality, according to the UX studies.

Meanwhile, experiential behaviors have the following attributes: intrinsic, ongoing search, and impulse buys. Hedonic quality may be experienced primarily in these behaviors. Hassenzahl et al. (2002) also argued that the users' evaluations varied depending on whether they had a goal to achieve or not. Hassenzahl (2004) addressed two human needs related to product context: competence/personal growth and relatedness/self-expression. He showed that when people interact with a product,

they have two different goal categories—do-goals and be-goals (the corresponding modes are goal and action modes). The do-goal mode refers to practical and task-oriented goals, and the be-goal mode refers to the behavior itself, and emphasizes the fun and entertainment experience obtained from the product. The consequences of user experience such as overall judgements, usage behavior and choice between alternatives are all outcomes user reaction as well as instrumental and non-instrumental quality perceptions. They regarded the emotional reaction as a component integrating instrumental and non-instrumental quality, not as a result.

Since Hassenzahl's UX framework, UX models have emerged that extend other variables surrounding user-product interactions. Figure 2.1.3 below is the UX model claimed by Helander and Khalid (2006), which consists of designer's environment and affective user experience.

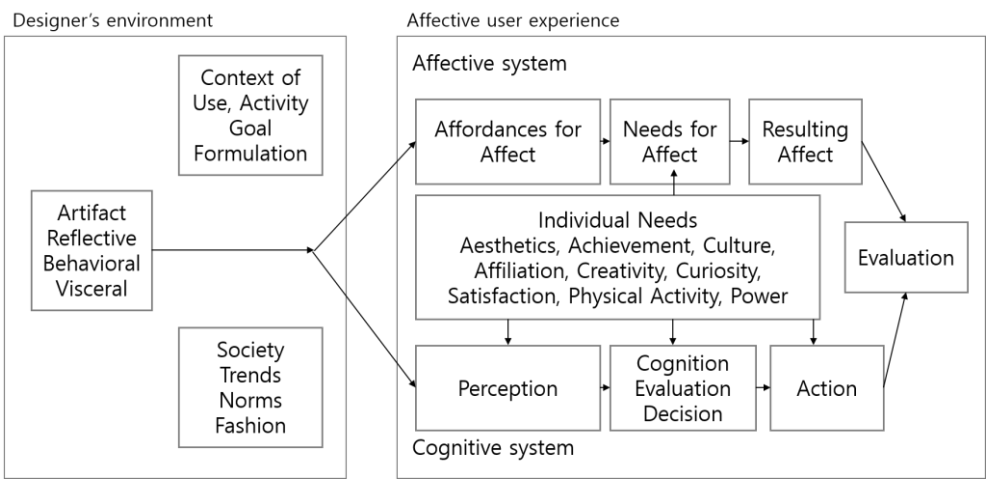


Figure 2.1.3 Affective user experience framework (reprint from Helander and Khalid, 2006)

Helander and Khalid presented the UX framework using two areas: designer perspective and user perspective. In the designer perspective, product characters,

context, and social attributes are the main variables. It is the stage where interactions do not occur, and no actual experience is achieved. In this area, only objective attributes, such as product attributes, trends, customs, and usages exist, and there is no subjective judgment that the user gains through the experience. The user perspective is the stage where the user actually interacts with the product and (perhaps) perceives and evaluates the product through interaction. Users' individual needs vary with aesthetics, achievement, culture, affiliation, and so on. Users experience and evaluate objects based on these needs.

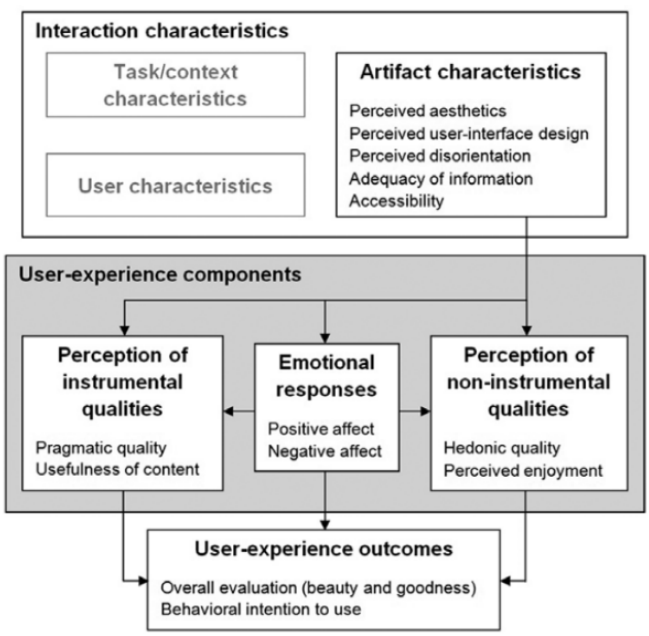


Figure 2.1.4 High-level representation of the UX model with news sites (reprint from Aranyi, 2015; based on Mahlke and Thüring (2007)'s UX model)

Aranyi (2015) suggested a UX framework that consists of interaction characteristics, user experience components and user experience outcomes (see Figure 2.1.4). This is a comprehensive UX framework based on the UX model proposed by Mahlke and

Thüring (2007). It is characterized by further expanding the detailed attributes and suggesting evaluation methods of components. Interaction characteristics consist of task and context characteristics, user characteristics and artifact characteristics (e.g. perceived aesthetics) and indicate the product character that exists before user-product interaction. The user-experience component is a component of perception of instrumental (e.g. pragmatic quality) and non-instrumental qualities (e.g. hedonic quality) and emotional responses (e.g. positive and negative affects). And the like. The perceived UX qualities and emotional responses obtained in this process lead to user experience outcomes. Through this process, the user has overall evaluation and behavioral intention to use.

2.1.2 The moments of UX

In addition to figure out the main factors affecting user experience, it is necessary to define how user experience is formed because user experience is closely related to the user satisfaction and continuous usage intention new technology. The phase of user experience can be divided into before UX, during UX and after UX. The before UX is the phase that the user does not interact with a product and the actual experience is not occur yet. Before UX is a step in perceiving a product's objective (or own) attributes, which refers to the pre-experience stage, which previous experience or prejudice affects. During UX is indicating the moments of actual "experiencing." As users interact with the product, an actual experience is generated. In this case, the usage experience is largely generated in two dimensions—pragmatic quality (PQ) and hedonic quality (HQ)—and the quality of experience in each of these dimensions is combined to generate the overall experience in the next stage, after UX. The after UX is the stage that an evaluation of the system is created based on previous experience. At this stage, the overall judgment is made based on

previous experiences through cognitive processes, such as means, self-images, memories, and messages. The concept of phase-based UX is linked to Norman's (2004) theory regarding levels of emotional design. Norman argued that the emotional design process consists of visceral, behavioral, and respective levels. The phase of before UX is equivalent to the visceral level. At this stage, the sensory quality (primarily visual quality) of the object and prior knowledge or previous experience with the product are indirectly affected. The user has a first impression of the product through a sensory encounter. The behavioral level is the stage where actual interaction takes place, which means during UX. The reflective levels include the state in which experiences are accumulated immediately after interactions have been made, as well as through repeated and continued use.

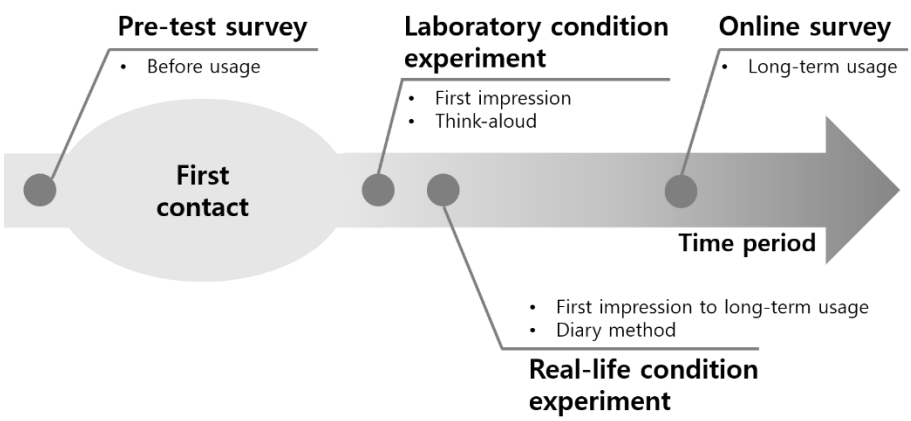


Figure 2.1.5 The phase based user experience process on the interactive products (Helander, 2002, reprinted)

Helander (2002) presented appropriate research methods for each stage of UX in empirical study (see Figure 2.1.5). In the first contact, a pre-test survey method is

used to collect impressions before the product is used. Next, the impression is examined in a think-aloud manner through a laboratory condition experiment. A real-life-condition experiment is used to obtain the experience data from using the system in everyday life. It is possible to check whether the experience has changed by comparing the first impression with the experience of long-term usage. Online surveys are also conducted to obtain insights on long-term usage.

Prolonged UX

Another variable to consider when constructing the UX framework is temporality (Kujala et al., 2000; Kujala et al., 2011; Mahlke & Thüring, 2007). Temporality is related to when or for how long the user experience is measured. Most UX-related studies measure the experience that occurs immediately after an interaction. Hassenzahl (2008) defines UX as "… a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service". According to Bargas-Avila and Hornbæk' (2011), 70% of UX-related user research studies were performed after UX, while only 17% of the studies analyzed all stages of UX. In some cases, it may be necessary to obtain experience data from short-term, fractional assessments, while long-term experience studies may be required in other cases. Dynamic aspects– on a small scale from minutes to hours as well as on a larger scale from days to months – of the user experience may be one of major trends for future user experience research (Mahlke, 2008). In other words, the scope of UX research includes a complex and wide range surrounding a product and a user. Adequate measurements should also be selected according to the purpose and necessity.

Recent research has primarily included long-term studies to measure UX (Walsh et al., 2014; Kujala et al., 2011; Luojus, 2012). In particular, everyday products that are used daily or frequently in everyday life show cumulative or changing UX with

repeated use. The experience data for everyday products may not be retained after the product is used once; rather, the experience may be enhanced or changed over time. Karapanos et al. (2009) recorded long-term observations on temporality showing how the quality of UX develops over time. They observed data from five weeks of use, from the moment the user purchased an Apple smartphone, and they found that extended use of the smartphone was triggered by a different quality beyond the initial positive experience. While early experience is primarily associated with the hedonic aspects of product use, ongoing experience is linked to how the product is meaningful to the user. The results of this study suggest that design keywords for meaningful mediation, designing for daily rituals, and designing for the self are important keywords for Computer – Human Interface (CHI) practice. Jang et al. From five years of self-reported diaries collected from TV viewers, it was observed that several of them mentioned the objective characteristics of products, such as content diversity, in the early stage of use. The frequency of mentioning satisfaction also increased over time.

According to Norman (2004) and Kujala et al. (2011), emotional quality is enhanced over time. Long-term studies show that the hedonic quality of the product (especially beauty) and usability during use are important in the early stages, and whether the product is related to the quality of life of the user becomes important during continued use. The UX process over time differs from Hassenzahl's UX model in that it describes the UX process from a more macro perspective, rather than looking at which experience factors affect each phase. It is important to note that the UX is not formed once, but rather cumulatively, through repeated use.

2.1.3 The UX framework with user experience, satisfaction and continual

usage intention

Based on the UX affecting factors and UX phases which are examined so far, it is also necessary to be defined how user experience affect user satisfaction and continual usage intention. User satisfaction exerts a significant effect on whether a technology is used continuously (Schulze and Krömker, 2010). Melone (1990) argued that user satisfaction is a certain type of evaluative response to information technology (IT). According to Hassenzahl's two-dimensional view of IT user experience (2003), perceived pragmatic performance and hedonic performance specify two primary evaluative dimensions of IT. This view is linked to the results of technology acceptance research. Perceived pragmatic performance is a strong predictor of the intended use of the technology (Davis, 1989; Venkatesh, 2000), and perceived pleasure is a measure of hedonic performance (Davis et al., 2013; Venkatesh, 2000). There are ample evidences that emotional responses through user-system interactions affect continuous intention to use (Kyun et al., 2009; Babin and Babin, 2001; Desmet and Hekkert, 2007).

Deng et al. (2006) asserted that user satisfaction affects the continuous usage intention and are influenced by pragmatic and hedonic performance, expectation disconfirmation and cognitive absorption. Expectation disconfirmation indicates the distance between expected satisfaction and experienced satisfaction. According to the expectation disconfirmation theory (EDT), which explains consumers' satisfaction with products (Spreng &Page, 2003; Bhattacherjee & Premkumar, 2004; Oliver, 1980), the disconfirmation of expectations specifies the nature of its impact on satisfaction. In other words, positive disconfirmation occurs when the perceived value is higher than expected. The higher the positive disconfirmation, the greater the satisfaction (Deng et al., 2006; Hsu et al., 2007).

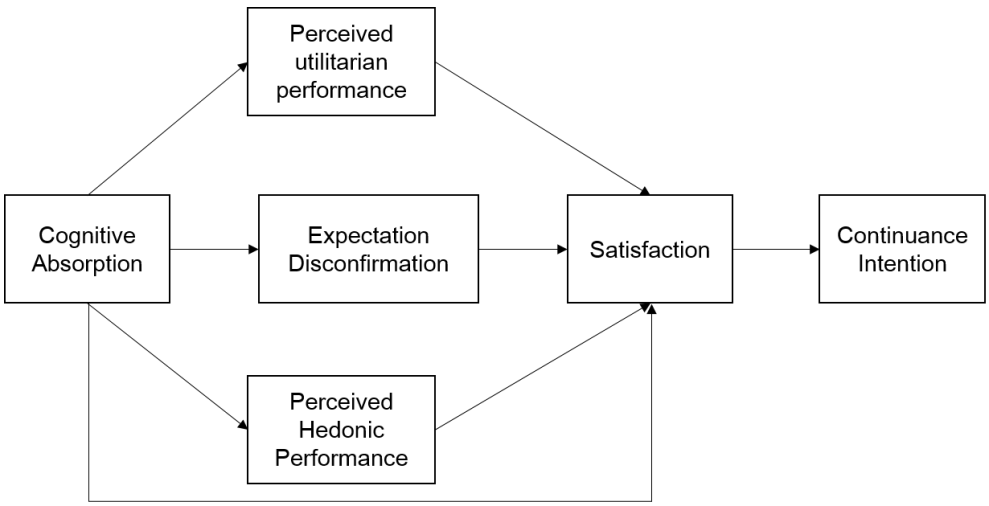


Figure 2.1.6 Research model of IT satisfaction and continual usage intention (Deng et al., 2006, reprinted)

2.2 The interactive product UX framework

In the relationship between user experience (UX) and user satisfaction, various hypotheses about the sub-factors of user experience that affect user satisfaction have been proposed. The technology acceptance model (TAM) proposed by Davis (1989) explains two primary predictors, perceived ease of use and perceived usefulness, of the potential user's behavioural intention. Research on e-commerce (Flavián et al., 2006), mobile phones (Sonderegger and Sauer, 2010), and mobile internet service (Amin et al., 2014) has shown that user satisfaction is the main factor contributing to intention to use and is associated with perceived ease of use and perceived usefulness. In addition, studies on information quality (DeLone and McLean, 1992; Lin and Lu, 2000) and system quality (Beyah et al., 2003; Cheong and Park, 2005) indicate that these factors are important predictors of ease of use and perceived

usefulness. These usability-related attributes are required attributes that help the user achieve his/her goals. In the field of human-computer interaction, user satisfaction is mainly derived from usability. This approach has contributed to enhancing functional quality, such as improving the usability of a product or increasing its efficiency. Improvements in usability are very effective in increasing user satisfaction, particularly when they are productive or supports users in achieving their goals.

However, TAM concentrate in the early phases of UX lifecycle – in the before use or in the first times of interaction (Roto, 2007). In contrast, user experience based research model starts from the first interaction phase to the continual usage intention as long as the user remembers the product. In addition, the user experience directly affects user satisfaction. Ogara et al. (2014) asserted that user experience and perceived richness are important drivers for user satisfaction. Ryu et al. (2008) argued that customer satisfaction are significant predictors of customers' behavioral intentions.

User satisfaction based on experience is an antecedent of continual usage intention of new technology. It has a greater meaning, especially in studies of everyday products such as TV. Interactive TV is required fun and relaxation rather than usability. Thus, when new technologies are incorporated into these everyday products, user experience based approach can be an adequate predictor whether the new technology is used continuously by users. Roto argued that UX research should be "worth-centered design" (Cockton, 2006) that includes the value aspect.

Based on the theoretical backgrounds so far, the following UX research model - the experiences (pragmatic and hedonic) of users with a new technology become an antecedent of user satisfaction and further intention to use the new technology - is

possible.

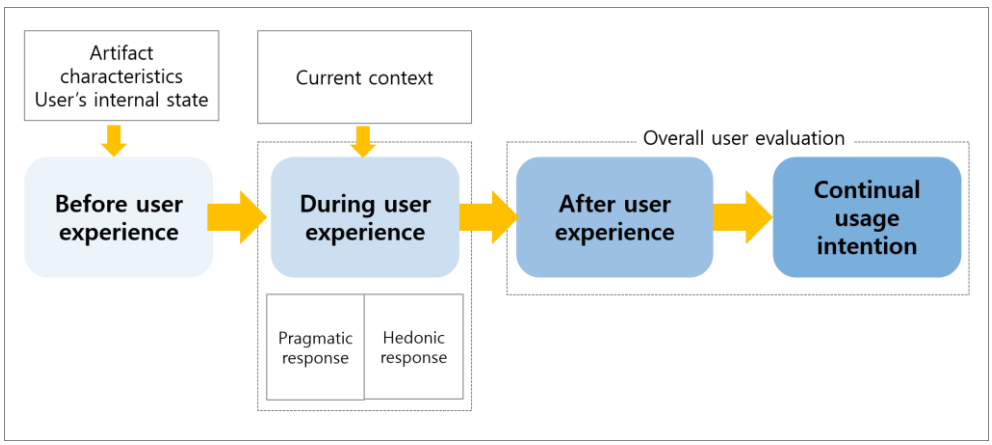


Figure 2.2.1 Integrated UX frameworks for interactive products

A research model which is developed to investigate how user experience influences user's satisfaction and continual use of IT are illustrated in Figure 2.2.1. This UX research model can provide a valuable insight into why and how the user forms positive experience of technology, as well as information on user satisfaction and continued use of new technology. There is a temporal relationship between each stage. Expectation occurs in the before UX stage that refers to the antecedent's stage before interaction occurs. At this stage, there is no direct experience with the product, so objective information-prior knowledge of the product, external prejudice, or external evaluation of the product character is made. During UX is the stage where a user perceives or is affected by an actual product interaction. At this stage, the user has perception of the product in two dimensions: perceived pragmatic qualities and perceived hedonic qualities. The user then evaluates the product based on the product qualities perceived through interaction in the after UX phase (Hassenzahl, 2001).

This user evaluation indicate overall user satisfaction and user satisfaction is linked

to the continual usage intention (Belanche et al., 2012). This UX cycle refers to a series of UX formation processes that evaluate a product/service through its use, from the first impression of the product. Some products have only a single use, but most products create a user experience through repeated use, rather than single use. In particular, most of the everyday products around us now assume daily use. Therefore, when measuring the experience of these products, it is important to study the experience of repeated use rather than one-time. This is because experiences gained from previous use can accumulate through repeated use, thereby intensifying or diminishing experiences by influencing already made positive or negative experiences.

2.3 Methods for evaluating user experience

Maja and Furtado (2016) analyzed 25 UX-related studies and pointed out the following problems: (1) psychophysiological research methods are not yet widely used for UX evaluation; (2) many researchers prefer qualitative studies than quantitative studies; (3) UX evaluations are mostly conducted one time and targeted already existing products; (4) mostly only one method is used, and do not consider prolonged use. In this paper, both subjective and objective methods were applied to explore interactive TV UX. This combined method aims to reduce controversies regarding reliability and validity that may arise when only one methodology is applied. Additionally, a richer interpretation of the results can be obtained by comparing the results of the two methods. There is ample evidence that the combination of psychophysiological and psychometric methods provides reliable measurements of affective UX (Maja and Furtado, 2016; Bengs et al., 2009; Ganglbauer et al., 2009; Vermeeren et al., 2010; Roto et al., 2009; Nacke et al., 2010)

2.3.1 Subjective measures

Because the UX comprehensively covers the subjective experience and satisfaction of the user, the user experience measurement has been mainly used for subjective measures such as self-report questionnaires, think-aloud and interviews. Many studies asserted that the user "emotion" plays an important role in the UX of entertainment products; however, it is very difficult to measure emotions and quantify them. Emotion data are usually obtained from results of subjective responses (self-report questionnaires), thinking aloud, heuristic evaluation, and contextual inquiry. Because it relies on the user's recall or subjective expression, which makes it difficult to measure objectively and focuses on the result rather than the process (Wilson & Sasse, 2000b). The emotional assessment should also consider a quantitative data measurement method, such as usability testing, which uses data based on an objective standard such as performance time or error frequency. In order to overcome these limitations, more studies are being applied to UX measurement, or subjective and objective measurement methods are combined. Mahlke (2008) asserted that a combination of methods that assess different components of user's emotional reactions provide a comprehensive basis for analyzing emotions in human-computer interaction.

Attrakdiff²

According to Hassenzahl, Diefenbach & Göritz (2010), a product's character is a set of product attributes that are created by a combination of product features and user expectation or standard. A product's character can be explained mainly by the following two attributes: pragmatic attributes (e.g., efficiency) and hedonic attributes (e.g., pleasure). Pragmatic attributes include usability factors such as usefulness and effectiveness, and hedonic (= emotional) attributes indicate emotional factors such

as fun and boredom. Hassenzahl (2004) proposed the AttrakDiff2 questionnaires to measure the pragmatic attributes and hedonic attributes. AttrakDiff2 consists of four quality dimensions, namely pragmatic quality (PQ), hedonic quality (HQ)-satisfaction (HQ-S), hedonic quality-identification (HQ-I), and APPEAL. It indicates the overall experience satisfaction and has a method of choosing a level similar to a person's own mood state by having mutually opposing (semantic differential) adjectives on both sides. The Semantic differential is an application of Osgood's method (1975) that selects words that are closest to the user's mood among opposing adjectives. After finishing each experiment, the participants evaluate subjectively the AttrakDiff^2 questions that ask about their own feelings. The PQ consists of seven items for the usefulness aspect such as "complicated–simple" and "unpredictable–predictable." The HQ consists of seven HQ-S items and seven HQ-I items. The HQ-S consists of "conventional–inventive," "dull–captivating," etc., and the HQ-I consists of emotional adjectives such as "isolating–connective" and "tacky–stylish." APPEAL refers to the overall UX, which combines the perceived HQ and PQ, and consists of seven items such as "repelling–appealing" and "bad–good." Each item consists of seven scales from −3 to +3, and a higher value indicates that the user has a more positive experience.

"Appeal" or "attractiveness" used in Hassenzahl's UX research model mean "the overall evaluation of users after use." Thus, "attractiveness," as used in this paper, refers to a more comprehensive user experience compared to the adjective meaning of "attractive." Examples of using the word "attractiveness" for overall evaluation are also shown in the studies of Isleifsdottir (2008) and Lindgaared and Whitfield (2004). However, when explaining the structural equation modelling of UX, the word "overall evaluation" is used instead of "attractiveness" to clarify the causality of the main factors of interactive TV UX.

Experience Sampling Method (ESM)

ESM (Larson and Csikszentmihalyi, 1983) is one of the daily diary methods and is called as ecological momentary assessment (EMA). This is an intensive longitudinal research methodology that involves asking participants to report on their thoughts, feelings, behaviors, and/or environment on multiple occasions over time (Wikipedia). This method consists of a measurement scale and open questions for the user to check and record daily. Participants record their status - thoughts, feelings, and behaviors - and respond to relevant questions, based on a specific time or signal every day.

2.3.2 Objective measures

A commonly used method of measuring usability is an objective measure of performance (e.g. performance measures), which involves the collection of quantitative data such as task complete time and errors. In the HCI field, it has been used to improve usability and product efficiency based on measurement results such as performance time and error rate. Effectiveness, efficiency and satisfaction (ISO 9241; ISO 25010) have been widely adopted in usability research and practice in HCI fields. These metrics are measurable and showed objective results. These methods are necessary to secure usability and usefulness, but there are limitations in UX research that sees what happens in real situations such as variables and relationships over various situations or time. Accordingly, methods that can obtain the objective measurement results while applying the actual situation are being applied.

Facial expressions

From Darwin's evolutionary theory (1872) to Ekman (1970), facial behaviors are associated with emotions. Ekman (1982) showed a method measuring facial expressions. According to Ekman et al. (), facial expressions imply emotion, cognitive processes, and social interaction. This method calculates the change of facial expression according to emotion based on the measurement values such as brow raising, brow lowing, lid raising, cheek raising, and lid tightening, etc. Predict.

Physiological measurements

One of the objective methods is physiological measurement (Kreibig, 2010; Mandryk, Inkpen & Calvert, 2006; Yao et al., 2014). Physiological measurement is a method generally used to determine human emotions in the psychology area, and includes heart rate (HR), skin conductance (SC), and facial muscle activity measurements. It involves the measurement of cognitive effort or emotional response by attaching a sensor on the face or the body and sensing the physiological state (frowning, smiling, perspiration, etc.) signals appearing as a result of interaction (Ekman, 1999). Levenson (1992) and Hamdi et al. (2012) asserted that physiological data sources clearly produce different results from each other depending on the emotions. In a physiological measurement experiment targeting game users, Mandryk, Inkpen and Calvert (2006) presented a result that the physical responses between games having a computer as opponent and games having a friend as opponent were different. They determined that a more specific and an accurate evaluation would be possible through a cross-verification of quantitative data and subjective data obtained directly from users. Yao et al. (2014) also collected physiological data and self-report data together in a UX evaluation of a mobile application. They asserted that through a result showing correlations between SC results and self-report data, physiological measurement can be one of the effective

methods for UX evaluation.

2.3.3 Summary

As with other studies, the criteria for selecting measurement methods in UX studies are dependent on the research goals, data required, and methods appropriate for obtaining that data. In particular, UX has a wide range of applications. However, it also has very diverse viewpoints such that it is dangerous to define the "overall experience" of a product or technology based on the result of a single measurement. Helander and Khalid (2007) suggested several possible methods while referring to the possibility of errors when applying only a single method. Roto et al. (2010) also insisted that although generally accepted overall measurements of UX do not exist, it can however be assessed in many different ways. In addition, many researchers have reported that applying one method in UX research may lead to misinterpretation, and hence, insisted on mixed-method approaches (Law, 2011; Arhippainen et al. al., 2013; van Turnhout et al., 2014).

In UX research, subjective measurements are used, in which users evaluate their experiences themselves. More recently, there are an increasing number of cases where objective methods such as physiological measurements are applied. Physiological measurements can measure the user experiences during user–system interaction without disturbing the user, and it is possible to secure objective and direct data. However, the results may not be clearly displayed sometimes. As a result, various attempts such as applying subjective measurements and objective measurements in parallel, or repeatedly measuring with a time difference, have been initiated. The application of these mixed methods can complement the limits by applying a single method or increase the reliability of the results through correlation between the results of both methods.

Another reason why the composite method needs to be applied is for the performance and preference relationship. In the HCI study, the higher the satisfaction of the usability, the higher was the preference for the product used. However, various studies have shown that the usability and preferences are different. To minimize errors by applying random measurements, various methods of cross-application or qualitative methods as well as quantitative measurement methods are required. This is necessary not only for theoretical development of UX research but also for practical aspects to define various approaches of UX-related methods as well as the strengths, weaknesses, and roles of each method.

Chapter 3
Interactive TV Studies

3.1 Definition of Interactive TV

TV is one of the typical entertainment products that provide pleasure at low cost. As the TV platform has been combined with the Internet and various functions (e.g., games and social media) can be accommodated, TV is evolving from mass media to interactive media. In this chapter, we define interactive TV and look at previous studies on it. The main contents of this chapter are (1) basic concept of TV and interactive TV definition, (2) interactive TV interface study, and (3) CA interface study which is combined with TV.

3.1.1 Traditional TV vs. Interactive TV

Before defining interactive TV, traditional TV should be defined. In terms of technology, traditional TV has the characteristics of being standardized, time-dependent, one-way, and public. While traditional TV performs one-way communication with an unspecified number of people, interactive TV does not. In terms of context, interactive TV is not limited by time and place, whereas traditional TV viewers gather in one place to watch a TV. Traditional TV has two main constraints - time and space. Time constraints indicate TV viewers must adjust to the

time provided by the content provider (e.g. the 9 o'clock news should be seen at 9 o'clock). Space constraints mean TV viewers should consume content where the TV is located such as in living room or bedroom.

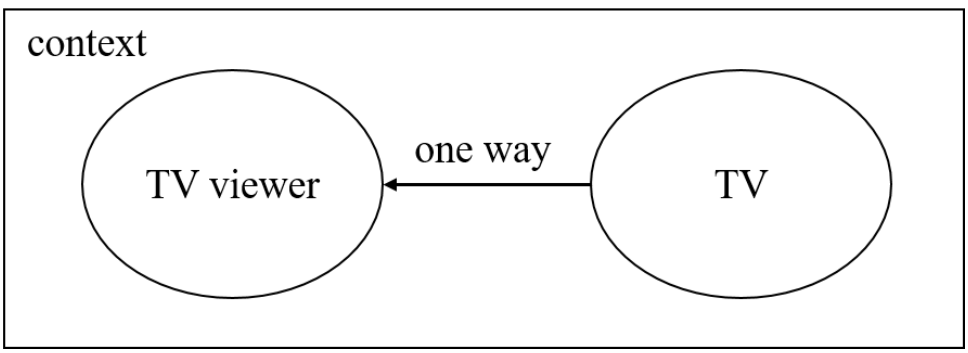


Figure 3.1.1 TV-viewer one way interaction

Interactive TV doesn’t have any time or space constraints. TV viewers can watch the TV what they want anytime and anywhere. Interactive TV began in the late 2000s. The development of the internet and technological developments have brought about internet protocol television (IPTV) and over the top (OTT) services. Since then, TV services and systems have appeared in the market.

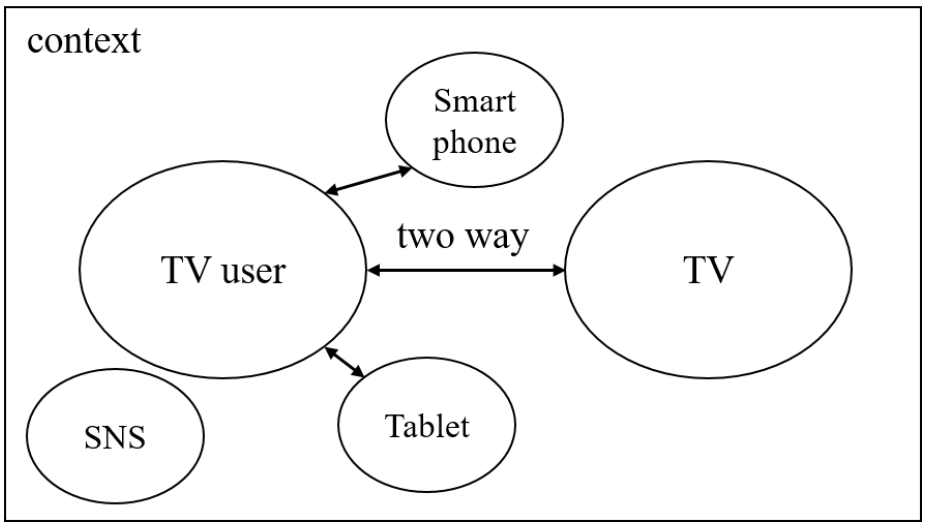


Figure 3.1.2 TV-viewer two way interaction

Connected TV (CTV) refers to any TV that can be connected to the internet and access content beyond what is available via the normal offering from a cable provider. It also refers to devices that use a television as a display and can connect to the internet to access content. Connected TV is also called Smart TV or Internet Protocol TV (IPTV) (Wikipedia). If interactive TV is defined according to the communication between the TV and the viwer, "enhanced TV" is a definition based on content. “Enhanced television can be defined as any interactive television service that makes an existing television programme better, while that programme is running and shortly afterwards. Enhanced television service providers typically add overlays, text and graphics to programmes, so viewers can interact while they watch” (Gawlinski 2003, p. 17).

Enhanced TV has content constraints. TV viewers must watch the limited content provided by the content provider (e.g. the film to be aired in the "Weekend Movie" program is pre-determined by the broadcasting station and unilaterally delivered to

viewers). In addition to the contents provided by the broadcaster, enhanced TV allows TV viewers to see the various contents such as personal broadcasting or internet broadcasting. The ability to display live scores or player information on the screen while voting in real time or watching soccer games is an example of enhanced TV that displays the information provided on the screen.

3.1.2 Interactive TV characters

The shift in TV viewing as technology advances is not happening inside the TV, but rather outside the TV. In the past, there was a 1: 1 relationship between TV and viewers. Now, users get information including TV-related content through various devices other than TVs. TV viewing channels are shifting from static TV to mobile TV. Previously, there was one TV in a home, now one person watches one TV. TV is also used as a background medium (while watching TV, 67% responded they only watch TV in the 1990s, only 34% of the same question in the 2000s (Hasebrink, 2001)). Now, instead of obtaining additional information related to TV content through TV, people use other devices (smartphones, tablets, etc.) while watching TV.

One of the main reasons for not using the new features available on interactive TVs was that users reported that they did not need them and were not comfortable using them. 'Not necessary' means that it can meet the needs in places other than the TV, and 'being inconvenient' implies an uncomfortable interface (mostly the RCU). New interfaces, such as keyboards, motion or voice recognition, and joysticks have emerged for interactive TV features, but few of them can replace the RCU. The reason for this is that not all of them provide an optimized experience for watching

The main purposes of watching TV are 'entertainment' and 'relaxation'. The entertainment pointed out here is not that which requires active interaction, such as games, but rather passive enjoyment where narrative storytelling is provided through

TV without mental effort. Therefore, for many features provided through TV, e.g., news, shopping, stock trading, and social network services (SNS), which are not narrative contents, TV viewers use more optimized devices such as smartphones or tablets. In other words, using a smartphone is more suitable for searching for information, while searching for information or SNS through TV using the remote control, is inconvenient for users. Although TV is an interactive product, it is more suitable for choosing than searching. These properties will be retained even if more suitable tools are available for TV interaction in the future. The TV screen interface is also not suitable for TV interaction. The interface of interactive TV is very similar to the interface structure and functions used in PCs. Tree structure menus for lots of content, and text input searches may not be adequate for the TV UX. 'Not necessary' and 'uncomfortable use' are related to each other. The inconvenience of the TV interface is the biggest reason preventing active use of TV by viewers in the evolution from static TV into interactive TV.

Jang et al. (2016) derived contents diversity, perceived picture quality and connectivity as the major interactive TV UX factors through several studies. Kunert (2008) divided TV features into hardware and contents, and analysed that hardware category consists of subsets of controllability, layout and hardware. The important criteria in contents category is quality, diversity, and efficiency (see Table 3.1.1).

Table 3.1.1 iTV product features (Kunert, 2008)

Category	Subset	Definition	Example
Hardware	Controllability	TV controller	RCU, voice recognition
		Control way	
	Layout	Screen layout	Menu structure, popup
	Hardware quality	Picture quality	3D
		Sound quality	Surround sound

Contents	Diversity Efficiency	Contents diversity Navigation	Game, SNS

3.2 Users

3.2.1 TV users

TV users indicate whoever interacts with a TV application. He or she can be called a user, a viewer or a consumer depending on the perspective taken (Chorianopoulos 2004). Users are primary users and secondary users. The primary user is the person who directly manipulates and watches TV. A secondary user refers the person who operates TV but not the main viewer. In case the mother manipulates TV content to show to the baby, the baby becomes the primary user and the mother becomes the secondary user. Traditional TVs would mainly be placed in the living room, and TV viewers would mainly be family members. Recently, people have been able to watch TV through more than one device. There are various alternatives, such as smartphones, cable, and streaming services, which allow users to watch the TV content whenever and wherever they want.

3.2.2 User goals

Watching TV is mostly done during the leisure time (Kunert, 2008), two key behaviours of TV viewers are information processing and entertainment (Chorianopoulos, 2008). In this regard, Cesar and Chorianopoulos (2008) claimed that the new TV experience should be designed by considering the TV contexts, and that the viewers should “keep on enjoying strong storytelling.” Information processing and entertainment, the major activities involved in watching TV, are correlated. The user performs information processing for the purpose of

entertainment. However, information processing by TV viewers is different from that by PC or smartphone users. TV viewers' information processing is closer to relaxed exploration rather than information seeking (Zimmerman, 2004).

3.2.3 Context of use

The purpose of watching TV is not very different between traditional TV and interactive TV. In addition, people watch TV in order to obtain a topic of conversation and keep up with trends. Hsu and Weng (2007) classified interactive TV services and derived three representative user tasks around contents independent service: watching TV programs, browsing information and performing business transactions. In the past, watching unilateral broadcasting content was a common situation. Now, more and more viewers are looking at only the scenes that they want to see. These changes are clearly distinguished by age. According to a 2017 survey of the use of public broadcasting services by Ofcom in the UK, the average time spent watching TV for young people decreased and content consumption patterns moved from legacy TVs to online services such as Netflix and Amazon. On the other hand, watching time for those over age 65 increased. The average time spent watching TV was 344 minutes a day. Instead of watching live TV narratively, users in their 20s and 30s who are accustomed to using digital appliances, can choose to watch only the clips that they want, or watch them at the desired time through mobile or streaming services. They now multi-task while watching TV. TV is also becoming a background media that just turns on (Kunert, 2008).

3.3 TV interfaces

TV interfaces include screens, electronic program guides (EPGs), hardware, such as

set-top boxes, and controllers. In the case of a screen displaying content, the size of static TVs in the living room is around 27 inches or more, and the average distance between the viewer and the TV is around 2 meters.

Table 3.3.1 TV interface features

Interface	RCU	CA
Operation	Button click	Voice order
Recognition rate	99%	80 ~ 90%
Detailed operation	O	△
Semantic awareness	X	O

3.3.1 Remote Control Unit (RCU)

The remote control consists of a four-way button, an OK button, ten numeric buttons, and an option button for using other manufacturers or service functions of a TV service provider. The remote control is optimized for traditional TV use such as channel zapping and volume control, and is not suitable for interactive context (e.g. text input).

3.3.2 Beyond RCU

As for a new control method for interactive TV service, a controller that is optimized for watching TV is still available, such as providing a keyboard which is installed on the back of the remote control, an on-screen keyboard displayed on the TV, and motion or voice recognition. The reasons why it's hard to replace the remote control are, people are familiar with using the remote control, and other tools are still inconvenient to use.

3.3.3 Interactive TV add-on new features: conversational agent (CA)

Conversational agent (CA) is one of the key interaction modality that cannot be avoided in the future. Based on the voice recognition technology, CAs, such as the recently developed Apple Siri, Amazon Alexa, and Google Assistant, provide information including weather data and assist tasks in our daily lives such as playing music. Furthermore, by combining it with a device such as TV, the CA performs the role of an assistant that helps the user. It is obvious that without changing the TV viewing context, the CA assists a user who feels uncomfortable with the complex TV control function (e.g., text input and content search). CAs appeared in the late 2000s. Because of low voice recognition rate and insufficient database, the CA technology was advancing at a lower rate. In recent years, the CA functions have been improving at a faster pace through technological advancements such as machine learning and cloud-based big data accumulation.

Figure 3.3.1 Conversation agent from upper-left side to clockwise: (a) SKT Nugu, (b) Apple Homepod, (c) Sony, (d) Naver Clova, (e) Kakao Mini, (f) Amazon Echo, (g) Samsung Galaxy home, (h) Whawei, (i) Amazon Echo dot, (j) Google home)

The most popular smart speaker in the world today is Alexa, powered by Amazon Echo, with 1.37 million units sold from 2016 to 2018 (Strategy Analytics, 2019). It is linked with various applications through an open platform policy and the expansion of low-cost speakers. In addition to Amazon, competition over speech recognition technologies, such as Google, is intensifying, because speech recognition technology will be the main platform of next generation. However, as pointed out by researches on smart agents by Levesque (2017), Bushnell (2018), Murnane (2018), and the review on Amazon Alexa by Reeves (2017), it has not yet reached the ultimate target of being a "natural computing paradigm."

In the overseas market, including the United States, artificial intelligence speakers are most prevalent, whereas in Korea, the most popular products are combined with IPTV (Consumer Insight, 2018). According to the 2018 Mobile Communication Planning Research of Consumer Insight in South Korea (2018), CA embedded IPTV (KT Giga Ginnie) 39%, SKT Nugu 26%, Naver Clova 16%, and other applications 19%. According to the report about CA user satisfaction, the South Korean CA user satisfaction has not even reached 50% (only 49%). The number one dissatisfaction of CA users was "voice command is not delivered properly" (50% out of total dissatisfactions). In other words, the CAs that have been implemented so far still need functional improvements. Furthermore, the interaction type is not at a natural "conversation" level, but remains as one of the control methods that use "voice."

Table 3.3.2 Quality attributes of conversational agent

Category	Attributes	Reference
Performance	- Manipulation	- Thieltges (2016)
Effectiveness	- Avoid inappropriate	- Morrissey &

		utterances	Kirakowski (2013)
Efficiency	Functionality	- Accurate speech synthesis - Interprets commands accurately - Linguistic accuracy - Execute requested tasks - Facilitate transactions - General ease of use	- Kuligowska (2015) - Eeuwen (2017) - Wallace (2003) - Ramos (2017) - Eeuwen (2017)
Satisfaction	Humanity	- Convincing, satisfying and natural interaction - Able to respond to specific questions	- Morrissey & Kirakowski (2013)
	Affects	- Convey personality - Give conversational cues - Provide emotional info. - Entertain participants - Respond to moods of participants	- Morrissey & Kirakowski (2013) - Pauletto et al. (2013) - Ramos (2017) - Meira & Canuto (2015)
	Ethics & Behavior	- Protect and respect privacy - Sensitivity to safety and social concerns - Trustworthiness - Awareness of social context	- Eeuwen (2017) - Miner et al. (2002) - Herzum et al. (2002) - Vetter (2002)
	Accessibility	- Responds to social cues - Detect meaning or intent	- Morrissey & Kirakowski (2013)

		- Wilson et al. (2017)

Speech recognition rate corresponds to usefulness and efficiency. Analyzing research on conversational agents shows that there are many different quality attributes depending on the point of view. Guerin et al. (2018) selected empathy, formality, facing, vocabulary and linguistic alignment as criteria for evaluating virtually any type of conversational agent scenarios. Morrissey and Kirakowski (2013) presented performance, humanity, affects and accessibility as key elements for CA quality assessment. Radziwell and Benton (2017) classified previous studies on CA based on HCI claims of effectiveness, efficiency, and satisfaction. In a CA report, Microsoft research suggested smoothness and effective communication as key elements of CA design, suggesting that observations of emotions and personality should be preceded. In other words, by identifying the user's emotions (short-term, immediate: valance, arousal) and personality (long-term: dominance, friendliness) and applying the results to the CA design, the CA can respond appropriately to the user's needs.

Emotions and personality are essential for arousal, valance, dominance, social class, and so on. They can be defined as invisible properties, which can be defined by the user's response speed, speech volume, pitch, etc. In summary, for affective interaction between the user and the CA, (1) identify information about the user's speaking speed, voice tone, gesture, situation (e.g. whether it is destination-oriented, etc.), and (2) the agent should make a decision gesture and tone based on this identification. If necessary, cameras can be introduced to capture gestures or facial expressions. (3) This collected user data can be used to determine how strong the

user's need to achieve the task is. For example, if the user needs to complete a task, the agent prefers to respond briefly. Unexpected interruptions, such as new news, can give a negative impact on user mood. Conversely, the user response to this unexpected interruption can be used to determine how strongly the user wants to complete the task. Other studies about interviewed and observed real assistants are used to gather user information and build CA-related data (Ghosh and Pherwani, 2015). They observed actual assistants and interviewed them to derive personality and communication style. They also presented linguistic and speech characters for voice assistants through crowdsourcing and co-creation dialog. According to these research results, arousal affects a number of easily observed behaviors: speech speed, amplitude, the size and speed of gestures, facial expression or posture. Emotional valance is signaled by facial expression, pitch contour and rhythm of speech. Dominant personality generates characteristic rhythms and amplitude of speech, as well as assertive postures and gestures. Friendliness is demonstrated through facial expressions, speech prosody, gesture and posture.

These studies commonly pointed out whether the user personifies the target product. For example, the more CA reveals personality such as empathy or humility, the more positive the user is. What matters is that people expect not "how much are they similar to human?" but "how much are they characterized?" Ramos (2017), who studied Facebook chatbot, insisted that people expressed empathy during interaction with the emoji even though they knew that was not real. According to the "computers are social actors" paradigm proposed by Nass, Steuer, Tauber, et al. (1993), when a person interacts with a computer, he/she reacts as if dealing with a human being subconsciously. Because the CA has a personification attribute called "voice," there is high possibility of personifying the counterpart during interaction with the CA. According to a chatbot study by Hill, Ford and Farreras (2015), the conversation

time of a user and chatbot is shorter than the conversation time with a person, but longer than an instant message; moreover, the contents and quality of conversation are different. Based on an analysis of online reviews of Amazon Alexa users, Purington, Taft, Sannon, et al. (2017) asserted that as social interactions increase, the tendency to personify the CA increases, and when the personification tendency is high, the user satisfaction is also high.

3.4 Summary

In this chapter, research on interactive TV properties, previous studies, and the definition of a conversational agent, which is expected to be a key feature of the future, were discussed. For a long time, the main purpose of TV usage and the TV experience were not changed. The evolution of the technical attributes of TV and the optimization of user attributes to accept and use them are essential to improving the interactive TV experience. The relationship between TV and viewers may be maintained, but contexts surrounding TV, TV features, usage behaviors, and interaction will continue to change as technology advances.

The CA allows direct manipulation and can be anthropomorphic, while it is technically still less mature than the RCU. The CA is not suitable as an interaction tool to replace a remote controller yet. As TV evolves into interactive products and new technologies continue to be combined, new interfaces are inevitable. Therefore, it would be meaningful to see if the CA could meet these needs. Analyzing TV viewer experiences makes sense to see whether the CA can become the controller for a new interactive TV. Users always want to solve their needs and wants quickly, easily, conveniently, and pleasurably. Unused features that do not meet these criteria are either naturally removed or lost.

Chapter 4
Experiment 1. Contextual analysis of interactive TV

4.1 Introduction

The user experience is affected not only by the product's own attributes but also by the user's behavior. According to Apter's (1989) reversal theory, users are affected more pragmatically when they are unconsciously or consciously exposed to a goal-oriented situation. Conversely, if a behavior itself is the goal, the hedonic aspects become more crucial factors to the users. Many studies on consumer experiences in online environments indicate that there are two types of consumption behavior: goal-directed and experiential (Novak et al. 2003; Sánchez-Franco and Roldán 2005). According to Novak et al. (2003), these two categories of behavior have distinct different attributes in important consumer behavior issues such as involvement, search and decision making. Goal-directed behaviors have the following attributes: extrinsic, directed search, and planned purchases. Meanwhile, experiential behaviors have the following attributes: intrinsic, ongoing search, and impulse buys. Hassenzahl et al. (2002) found that the users' evaluations varied depending on whether they had a goal to achieve or not.

TV has been classified as a passive entertainment medium providing relaxation and enjoyment. Interactivity enables more active user participation. People may want to use interactive TV more extensively. On the contrary, people may prefer to just watch TV or may not like the interactive attributes of TV. The purpose of this study is to identify opportunities to see how interacting with TV in the different situations affects the TV viewer's experiences. To do this, I assumed that the match between an individual's motivational orientation and a particular product attribute (i.e., ergonomic or hedonic) would determine the perceived value of the interactive product, and I applied this to interactive TV viewers.

4.2 Method

This study has adopted Hassenzahl and his colleagues' user experience model (2000) to investigate how interactive TV viewers' experiences are differentiated in behavioral situations. According to Hassenzahl et al. (2000), a product is experienced to users through four different quality dimensions: objective quality (before interaction), perceived quality (during interaction), evaluation of the software (after interaction), and consequences of using the product. Users interact with the product and perceive the product's character differently depending on the behavioral situations. The product's character consists of two main attributes: ergonomic attributes (e.g., effectiveness) and hedonic attributes (e.g., fun and relaxation).

Before starting the main study, a survey was conducted to determine the main behavior of TV viewers. I asked participants about the main purpose of watching TV (N = ninety-nine: female = fifty-eight, male = forty-one, average age = thirty-six

years). I created a questionnaire about watching TV and administered the survey through SurveyMonkey.com (San Mateo, CA, USA), which is an online survey site. It was found that sixty-eight percent of the respondents habitually turn on the TV to pass time, or to find “something fun.” They search for a TV program by “channel zapping” or by navigating the menu. The average search time was three to five minutes. In addition, thirty-one percent said that they turn on the TV to watch a particular program. They preferred to reach the target content directly and quickly. The average time to reach the intended content was one to two minutes. The results showed that differences exist in the TV viewers’ behavior and interaction time based on whether they have a goal. TV viewers usually find interesting content through browsing or menu navigating, whereas when they have specific target content, they go directly to the destination.

In this experiment, the interactive TV service, which is provided by Korea Telecom (KT), was used as the material (Figure 4.2.1 below). The name of this interactive TV service is Olleh TV. Olleh TV is a market-leading Internet protocol TV service in Korea that provide thousands of live channels, video content, and applications. The Olleh TV service interface has a hierarchical menu structure suitable for use with four directional buttons on the remote control. Using a remote control, users navigate the menu in a vertical direction to first search for a category and then select sub-content belonging to the category in a horizontal direction.

Forty participants took part in this study (male = twenty-two, female = eighteen, mean age = twenty-nine years). All participants were new employees of KT. I recruited the participants from departments that were not directly related to the interactive TV service to avoid distorting the results of the experiment. All participants were familiar with the use of a remote control and knew how to search

for channels and find content. No participation fee was paid to them. To allow participants to watch and manipulate the TV comfortably, a separate room with a comfortable chair was used. The distance between the TV and participants was about two meters. A fifty-inch high definition TV was used in the experiment.

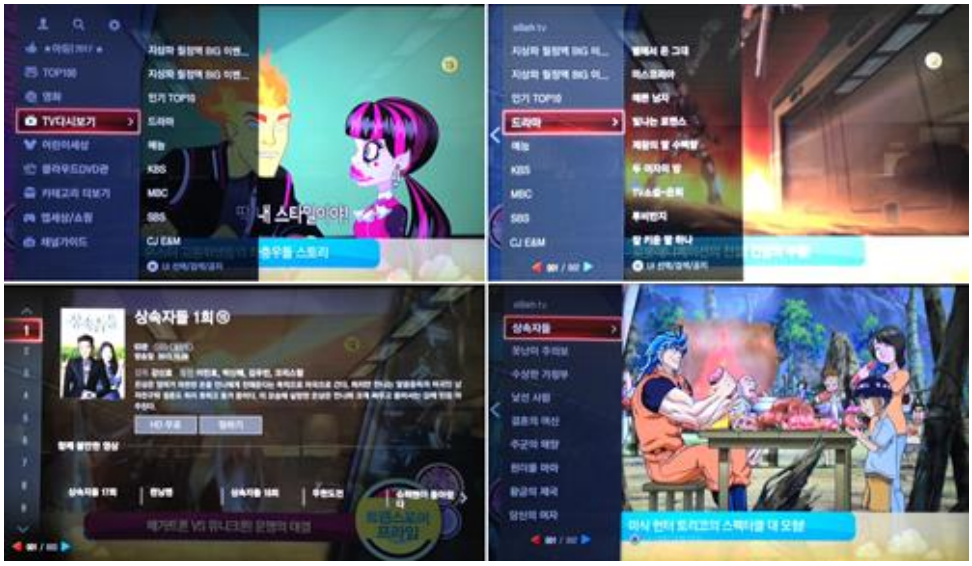


Figure 4.2.1 User interfaces of KT IPTV service (text-based UI)

Regarding the scale for measuring the Hedonic Quality (HQ), Pragmatic Quality (PQ), and Appeal (APPEAL) in terms of the user's experience, the Semantic Differential Questionnaire presented by Hassenzahl (2001) was employed. The Semantic Differential Questionnaire is a rating scale to measure the connotative meaning of products and concepts. HQ is composed of seven questions and consists of emotional adjectives such as "interesting - boring," "exciting - dull" and so on. PQ is composed of eight questions about usability aspects such as "comprehensible - incomprehensible" or "supporting - obstructing." APPEAL is the integrated user

feeling consisting of perceived HQ and PQ and has eight questions on dimensions such as “pleasant - unpleasant,” or “good - bad." Each item consists of 7 points from -3 to +3, which means that the larger the negative value, the more the users are experienced. All participants signed an agreement to participate before starting the experiment. Forty participants were divided into two groups. The first group performed the experiential tasks for ten minutes of freely navigating channels and finding content as they normally do when watching a TV. After using the TV, the participants completed the semantic differential questionnaire about the perceived quality of the product. In order to prevent errors in the experimental sequence, all the sequences were randomized. The second group performed three specific goal-directed tasks: (1) searching for a specific program’s title through menu navigation, (2) finding specific video-on-demand content, and (3) registering their preferred channels. Each task’s time was limited to three minutes in the goal-directed session. The experimental difficulty was adjusted appropriately so that all participants could complete the task in time. After completing the assignment, participants also answered the same questionnaires mentioned above. The task order was counterbalanced for the goal-directed group.

4.3 Results

A factor analysis was performed to determine whether TV viewers perceived pragmatic quality (PQ) and hedonic quality (HQ) independently in each task. A factor analysis for the experiential tasks extracted a main PQ factor and a main HQ factor from the semantic differential. The eigenvalue of PQ factor was 5.52 (36.81%) and HQ was 5.39 (35.96%) and the combined result explained 73% of the total variance (see Table 4.3.1). As a result of factor analysis on APPEAL, APPEAL

attributes were considered homogeneous because only one factor was extracted (eigenvalue = 6.29, explained variance = 78.62).

Table 4.3.1 The results of factor analysis for UX qualities in the experiential situation.

The user perceived product's attributes	Principal Components with Varimax	
	Factor 1	Factor 2
E_PQ1 comprehensible – incomprehensible	.850	.072
E_PQ2 supporting - obstructing	.785	.154
E_PQ3 simple – complex	.782	-.161
E_PQ4 predictable – unpredictable	.885	-.214
E_PQ5 clear – confusing	.886	0.1
E_PQ6 trustworthy – shady	.697	.078
E_PQ7 controllable - uncontrollable	.726	-.175
E_PQ8 familiar - strange	.804	-.142
E_HQ1 interesting – boring	.133	.789
E_HQ2 costly – cheap	.378	.809
E_HQ3 exciting – dull	.1	.877
E_HQ4 exclusive – standard	-.26	.905
E_HQ5 impressive – nondescript	-.252	.879
E_HQ6 original – ordinary	-.112	.885
E_HQ7 innovative - conservative	-.169	.898
Eigenvalue	5.52	5.39
% variance explained	36.814	35.963

Note. G, goal-directed; PQ, ergonomic quality; HQ, hedonic quality.

Table 4.3.2 Principal component analysis result for APPEAL in the experiential situation

The user perceived product's attributes	Factor 1
E_ APP1 pleasant-unpleasant	.851
E_ APP2 good-bad	.788
E_ APP3 aesthetic-unaesthetic	.891
E_ APP4 inviting-rejecting	.940
E_ APP5 attractive-unattractive	.917
E_ APP6 sympathetic-unsympathetic	.919
E_ APP7 motivating-discouraging	.843
E_ APP8 desirable-undesirable	.932
Eigenvalue	6.29
% variance explained	78.62

Note. APP, APPEAL.

The factor analysis extracted two PQ factors and a main HQ factor from the semantic differential for the goal-directed tasks. All the sub-attributes of HQ were similar to one another and the eigenvalue of the extracted factor was 6.39 (42.64%); the

explanatory range was 42.64% of the total variance. The eigenvalue of factor 1 extracted from PQ was 3.34 (22.26%) and factor 2 was 1.95 (13%); both factors accounted for 35.26% of the total variance. The combined PQ and HQ factors explained 78% of the total variance (see Table 4.3.3). Like the experiential session, only one factor was extracted from APPEAL (eigenvalue = 6.251, explained variance = 78.134).

Table 4.3.3 The results of factor analysis for UX qualities in the goal-directed situation.

The user perceived product's attributes	Principal Components with Varimax		
	Factor 1	Factor 2	Factor 3
G_PQ1 comprehensible – incomprehensible	.094	.840	.259
G_PQ2 supporting - obstructing	.516	.633	.149
G_PQ3 simple – complex	.25	.834	-.005
G_PQ4 predictable – unpredictable	-.068	.310	.852
G_PQ5 clear – confusing	.128	.544	.585
G_PQ6 trustworthy – shady	-.063	.030	.852
G_PQ7 controllable - uncontrollable	.074	.794	.202
G_PQ8 familiar - strange	-.424	.642	.025
G_HQ1 interesting – boring	.852	.171	-.101
G_HQ2 costly – cheap	.870	.03	.072
G_HQ3 exciting – dull	.919	.159	-.031
G_HQ4 exclusive – standard	.927	-.029	.003
G_HQ5 impressive – nondescript	.930	.163	.068
G_HQ6 original – ordinary	.969	-.039	-.053
G_HQ7 innovative - conservative	.924	.136	-.067
Eigenvalue	6.39	3.34	1.95
% variance explained	42.64	22.26	13

Note. G, goal-directed; PQ, ergonomic quality; HQ, hedonic quality.

Table 4.3.4 Principal component analysis result for APPEAL in the goal-directed situation

The user perceived product's attributes	Factor 1
G_ APP1 pleasant-unpleasant	.886
G_ APP2 good-bad	.859
G_ APP3 aesthetic-unaesthetic	.876
G_ APP4 inviting-rejecting	.906
G_ APP5 attractive-unattractive	.882
G_ APP6 sympathetic-unsympathetic	.866
G_ APP7 motivating-discouraging	.903
G_ APP8 desirable-undesirable	.891

Eigenvalue	6.251
% variance explained	78.134

Note. APP, APPEAL.

4.3.1 The comparison results between the experiential situations and the goal-directed situations

To test the hypothesis that the experience aspects users consider will differ between goal-directed situations and experiential situations, an independent samples t-test was performed. As can be seen in Table 4.3.5, PQ and HQ in the each situation (experiential situation and goal-directed situation) are significantly different from one another. There was significant differences in the scores for PQ ($t(39) = 2.454$, $p < .05$) in the experiential tasks and the goal-directed tasks. The results of HQ ($t(39) = 2.000$, $p > .05$) and APPEAL ($t(39) = 1.078$, $p > .05$) show no significant difference between the experiential tasks and the goal-directed tasks.

Table 4.3.5 The result of independent samples t-test

		Mean	S.D.	t	df	Sig. (2-tailed)
PQ	Goal-directed	1.36250	.89147	2.454	39	.019*
	Experiential	.79063	1.17615			
HQ	Goal-directed	.50714	.72913	2.000	39	.052
	Experiential	.08214	.50095			
APPEAL	Goal-directed	.91250	.59480	1.078	39	.288
	Experiential	.69375	.67768			

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

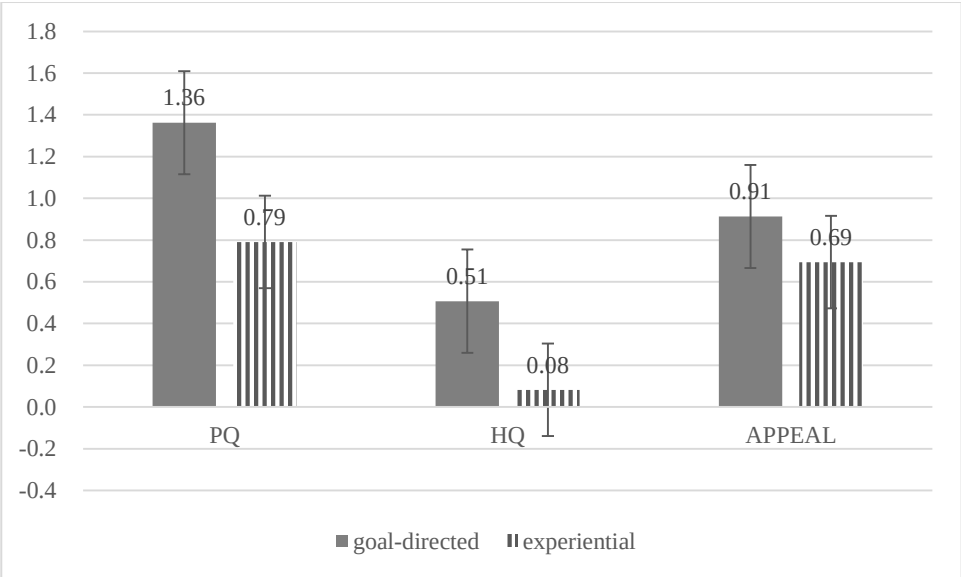


Figure 4.3.1 The mean values for PQ, HQ and APPEAL in the experiential situations and the goal-directed situations

Figure 4.3.1 shows the mean values among the UX quality variables for each situation. UX mean values of goal-directed sessions were more positive than experiential session mean values. All experience quality was more positively experienced in goal-directed sessions. In HQ and APPEAL, there were no significant difference between goal-directed and experiential situations. The mean value of APPEAL was located in between the PQ and the HQ that it can assume that the values of PQ and HQ indicate APPEAL, the overall experience.

Regression analysis was performed to investigate the effects of PQ and HQ on APPEAL in the experiential session. The result showed that both PQ and HQ affected APPEAL respectively and HQ had a little more influences than PQ on APPEAL. A linear regression was calculated to predict appeal of each interface based on UX sub dimensions (see Table 4.3.6). In the experiential session, a

significant regression equation was found (F (2, 37) = 63.103, p < .001), with an R^2 of 0030.773. As a result of confirming the significance of each dimension, it was confirmed that PQ (t(20) = 5.251, p < .01) and HQ (t(20) = 10.344, p < .01) had an effect on APPEAL. Both PQ (B = .393) and HQ (B = .716) were positive and the influence of HQ (β = .813) on APPEAL was twice that of PQ (β = .413). The explanatory power explaining the dependent variable by the independent variables was 77.3%.

Table 4.3.6 The results of ANOVA for evaluating the effects of PQ and HQ on APPEAL in the experiential session.

		R	R^2	R^2(adj)	B	Standard Error	β	t	sig.
Experiential session	PQ				.393	.075	.413	5.251	.000***
	HQ				.716	.069	.813	10.344	.000***
		.879	.773	.761					

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

Regression analysis was conducted to examine the effect of PQ and HQ on APPEAL in goal-directed session. The result showed that both PQ and HQ had a significant effect on APPEAL. In the goal-directed session, a significant regression equation was found (F (2, 37) = 81.535, p < .001), with an R^2 of .815. As a result of confirming the significance of each dimension, it was confirmed that PQ (t(20) = 5.942, p < .001) and HQ (t(20) = 10.333, p < .01) had an effect on APPEAL. Both PQ (B = .529) and HQ (B = .579) were positive and the influence of HQ (β = .738) on APPEAL was also much higher than PQ (β = .425). The explanatory power explaining the dependent variable by the independent variables was 81.5%.

Table 4.3.7 The results of ANOVA for evaluating the effects of PQ and HQ on APPEAL in the goal-directed session.

		R	R^2	R^2(adj)	B	Standard Error	β	t	sig.
Goal-directed session	PQ				.529	.089	.425	5.942	.000***
	HQ				.579	.056	.738	10.333	.000***
		.903	.815	.805					

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

Similarly, the correlation analysis results also showed a positive linear relationship between appeal (APP) and pragmatic quality (PQ), and appeal and hedonic quality (HQ). In Table 4.3.8, it was shown that the correlation analysis results between APPEAL and UX quality dimensions (PQ and HQ) in the experiential session (APP x PQ: r = .343*, APP x HQ: r = .777***) and the goal-directed session (APP x PQ: r = .531***, APP x HQ: r = .799**). The interaction between PQ and HQ was not shown.

Table 4.3.8 The comparison results of the effects of UX attributes on APPEAL between the experiential session and the goal-directed session.

	APPEAL		
	UX dimension	Pearson corr.	Sig. (2-tailed)
Experiential session	PQ	.343	.030*
	HQ	.777	.000***
Goal-directed session	PQ	.531	.000***
	HQ	.799	.000***

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

Figure 4.3.2 shows the result of the correlation analysis between APPEAL, PQ and

HQ in each situation.

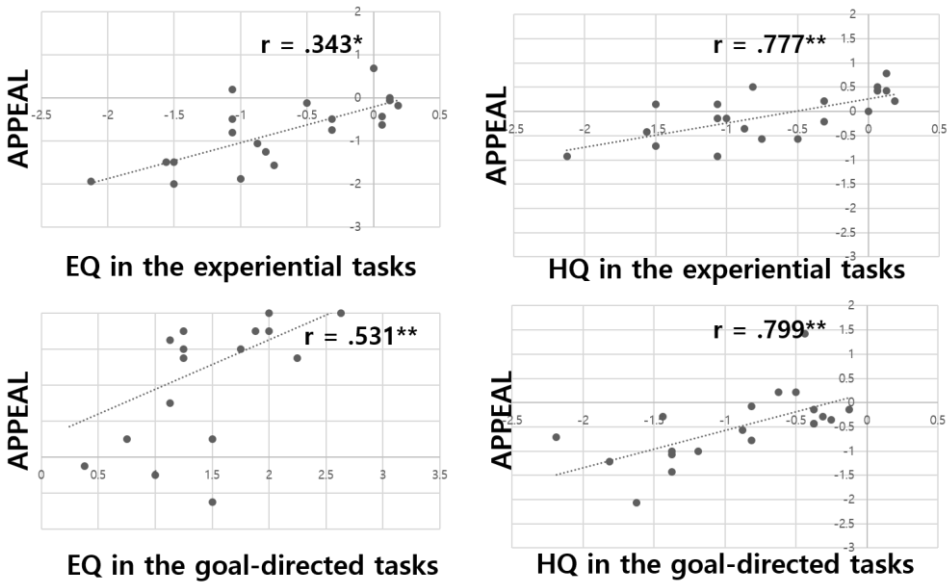


Figure 4.3.2 The results of correlational analysis

The following is the result of comparing the difference between experiential session and goal-directed session according to the detailed attributes of each quality variable (see Table 4.3.9)

Table 4.3.9 Paired t-test results for UX attributes in two contextual situations

	Attributes	Goal-directed mean (SD)	Experiential mean (SD)	t	sig. (2-tailed)
PQ	comprehensible – incomprehensible	1.60 (1.27)	.90 (1.52)	1.953	.058
	supporting - obstructing	1.28 (1.11)	.95 (1.38)	1.198	.238
	simple – complex	1.00 (1.43)	.28 (1.62)	2.159	.037*
	predictable – unpredictable	1.30 (1.44)	.83 (1.52)	1.633	.110

	clear – confusing	1.23 (1.35)	.73 (1.47)	1.669	.103
	trustworthy – shady	1.58 (1.06)	1.18 (1.11)	1.552	.129
	controllable - uncontrollable	1.60 (1.22)	.73 (1.60)	2.857	.007**
	familiar - strange	1.33 (1.44)	.75 (1.44)	1.733	.091
	interesting-boring	1.08 (1.54)	.63 (1.51)	1.623	.113
	costly-cheap	.43 (1.48)	.23 (1.42)	.732	.468
HQ	exciting-dull	.68 (1.42)	.40 (1.49)	1.076	.289
	exclusive-standard	.15 (1.56)	.08 (1.60)	.884	.382
	impressive-nondescript	.58 (1.48)	.05 (1.50)	2.139	.039*
	original-ordinary	.28 (1.69)	.40 (1.42)	2.683	.011*
	innovative-conservative	.38 (1.58)	.25 (1.31)	2.405	.021*
	pleasant-unpleasant	1.08 (1.07)	.83 (.98)	1.220	.230
	good-bad	1.38 (1.05)	.90 (1.01)	2.211	.033*
	aesthetic-unaesthetic	.55 (1.47)	1.00 (1.28)	1.584	.121
APP	inviting-rejecting	.78 (1.16)	.60 (1.17)	.827	.413
EAL	attractive-unattractive	.83 (1.31)	.45 (1.39)	1.391	.172
	sympathetic-unsympathetic	.93 (1.38)	.73 (1.24)	.813	.421
	motivating-discouraging	.88 (1.36)	.43 (1.49)	2.126	.040*
	desirable-undesirable	.90 (1.24)	.63 (1.55)	.968	.339

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

Figure 4.3.3 is about the result of comparing the difference between the two sessions according to the PQ sub-properties. All of the attributes included in the ergonomic quality dimension were more positive in the goal-directed situation than in the experiential situation. Significant differences between the two sessions are “simple – complex” ($t(40) = 2.159$, $p < .01$) and “controllable – uncontrollable” ($t(40) = 2.857$, $p < .001$).

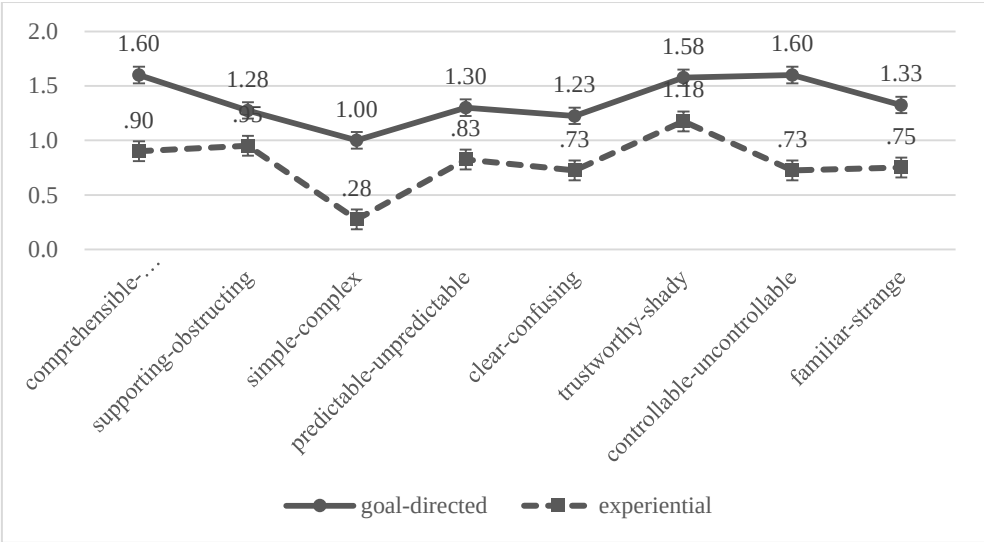


Figure 4.3.3 The comparison results of PQ between the experiential situation and the goal-directed situation

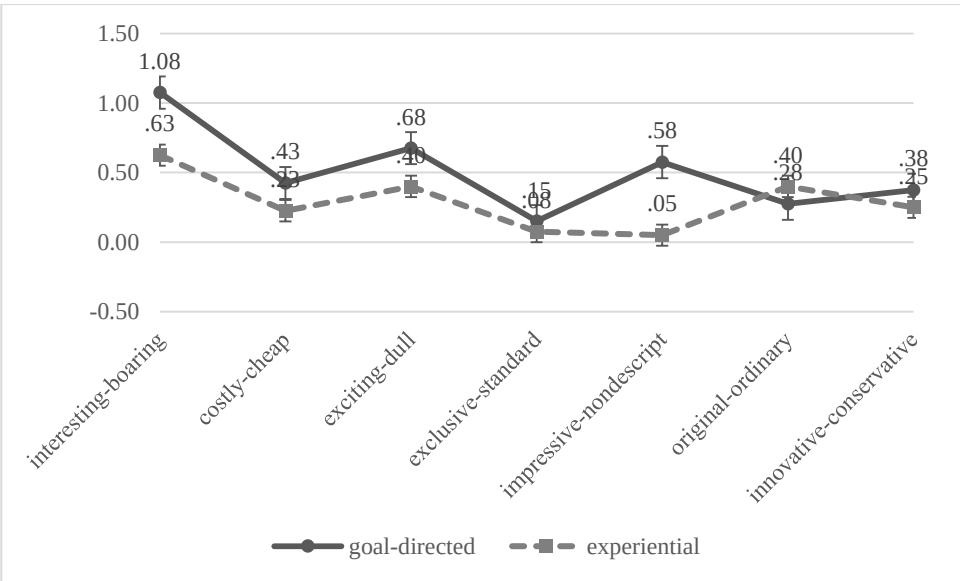


Figure 4.3.4 The comparison results of HQ between the experiential situation and the goal-directed situation

Figure 4.3.4 shows the comparison results of the two sessions according to the HQ sub-properties. All attribute values were higher in goal-directed situations than in experiential situations. The significant differences between the properties are "impressive-nondescript" ($t(40) = 2.139$, $p < .05$), "original-ordinary" ($t(40) = 2.683$, $p < .05$) and "innovative-conservative" ($t(40) = 2.405$, $p < .05$).

APPEAL also showed that all attribute values were higher in goal-directed situations than in experiential situations (see Figure 4.3.5 below). The significant differences between the properties are "good-bad" ($t(40) = 2.211$, $p < .05$) and "motivating-discouraging" ($t(40) = 2.216$, $p < .05$).

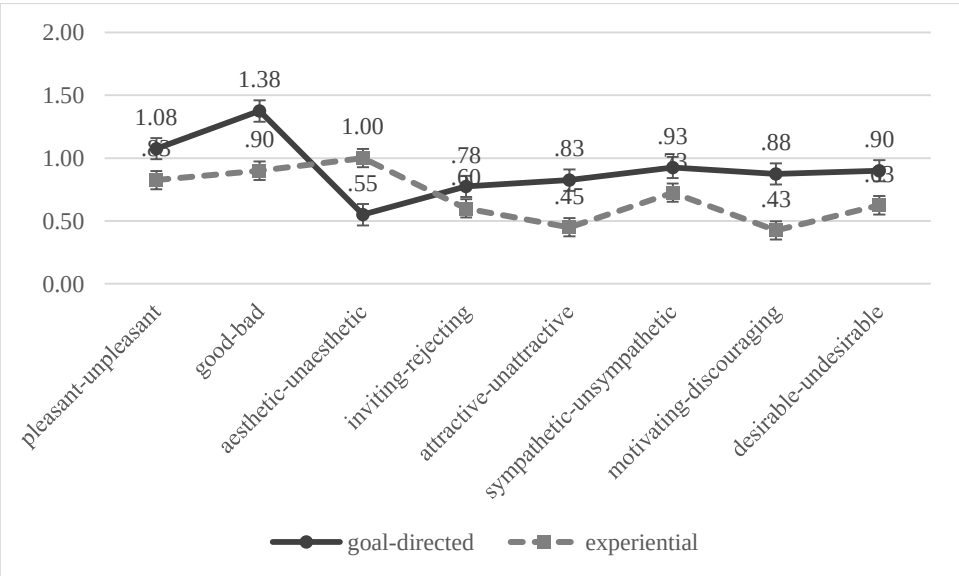


Figure 4.3.5 The comparison results of APPEAL between the experiential situation and the goal-directed situation

4.4 Conclusion

In this paper, how the TV viewers experience differs in the experiential situation and the goal-directed situation is investigated. First, it is confirmed that PQ and HQ are main factors of user experience measurements and independent each other, through factor analysis. Next, whether the experience of both sessions are different each other is measured. The results for the PQ, HQ, and APPEAL were significantly different according to whether participants were performing goal-directed tasks or experiential tasks. As a result of the evaluation of the two sessions, TV viewers were more experienced in the goal-directed situation. The result of regression analysis showed that both PQ and HQ had a positive effect on APPEAL. The similar results were shown in the correlation analysis that a positive linear relationship between APPEAL and PQ and HQ in the experiential tasks and the goal-directed tasks.

These results indicate again that the user perceive the product's character through ergonomic quality and the hedonic quality together regardless of the situation. Interestingly, the results of the regression analysis showed that the contribution of the HQ was almost twice as much as the PQ for the APPEAL in both situations. This implies the fundamental purpose of the user-TV interaction is to seek pleasure, and achieving the goal is also the fun seeking process. This also implies that hedonic aspect is the important factor than the ergonomic aspect on the TV-viewing experience. In this study, it was demonstrated that users had a different experience according to the different usage modes.

Also, the TV viewers in this study reported having a "better experience" when they could achieve the goals. The goals mentioned here are mainly about pleasure rather than productivity. This finding means that the process of achieving the entertainment goals can also be a part of pleasure, and interactive TV can provide a better

experience by providing various ways to experience pleasure corresponding to the user needs. As Vorderer (2001) mentioned, TV is not used for work but for entertainment, and research into the entertainment experience should examine ways of moderating the users' moods, focusing on their emotional outcomes.

Implication means that to improve the overall experience, a more positive experience can be derived when the specific task of achieving the objective is given from the user's point of view and providing a visually more beautiful system environment for the designer. It is necessary to determine how users' attitudes change based on their experience and the environment surrounding the TV, and to provide users with an optimized TV experience considering the changes in the TV. Good emotions create a favorable experience, and the more that people experience, the deeper the relationship with the medium.

Chapter 5
Experiment 2: Analysis of conversational agent-equipped TV UX

5.1 Introduction

This experiment is evaluated user experience through subjective and objective measurement methods. Self-report questionnaires was used as the subjective method which the user directly assessed. As an objective measurement method, physiological measurement was used to measure the feelings or the user's internal state during interaction.

5.2 Method

The experiment was conducted at the Neuro-Management Laboratory, Korea Advanced Institute of Science and Technology, for 10 days from November 26 to December 6, 2018. 42 people (21 males and 21 females) participated including the pilot participants. The experiment participants were students of the Institute, nearby university students, and office workers. Their ages ranged from 20 to 44 years and the average age was 29.02 years. The experiment was carried out in the following

sequence: (1) the experiment was explained to the participants first and then, the experiment participation agreement was signed; (2) sensors were attached for physiological response data collection; (3) the control method for interface 1 was studied and preliminarily evaluated; (4) the experiment was performed; (5) the experiment result was evaluated; (6) steps (3) to (5) were repeated for interface 2. The total time consumed was 1 h 10 min on average, and the sequence of RCU and CA controls was counterbalanced by a random method. In the preliminary evaluation before the experiment, the participants answered to questions regarding their expectations on the control method, and in the post-event evaluation conducted after each experiment, the participants responded on their experience after using it. That is, a subjective evaluation method was adopted. The physiological data of participants were stored separately through sensors attached on their faces and bodies during the experiments.

5.2.1 Tasks for TV UX

In the experiment, 14 different tasks related to TV viewings were performed (Table 5.2.1). According to the difficulty level, they were composed of 50% easy tasks, 30% average tasks, and 20% very difficult tasks. The performing sequence, performing time, goal, and difficulty level for the tasks in the CA session and those in the RCU session were designed to be as similar as possible. Task 8 is about controlling the video-on-demand (VOD) content. The task is complicated that I separated it into the two sub tasks (T8a and T8b). Reliability analysis results for the CA tasks and RCU tasks showed a relatively high internal consistency (Cronbach's alpha = 0.789).

Table 5.2.1 Tasks for the experiments of CA and RCU.

Task no.	Difficulty	Goal of task	CA Mean time(sec.)	RCU Mean time(sec.)
T1	Low	TV on	12.59	10.11
T2	Low	Volume up	5.2	2.49
T3	Low	Volume down	5.5	3.2
T4	Low/medium	Channel zapping to near channel	5.42	6.2
T5	Medium/high	Channel zapping to remote channel	4.93	18.4
T6	Medium	Menu navigation	6.29	6.96
T7	High	Search & play	20.52	29.8
T8a	Medium	VOD control	5.83	3.76
T8b	Low	VOD play off	7.9	3.39
T9	Low	Channel zapping	5.4	3.2
T10	High	Search & navigation	7.89	11.32
T11	Medium	Menu navigation	6.27	2.93
T12	Low	Volume control	5.53	2.59
T13	Low	TV off	5.39	1.9
		Total time	99.26	103.05
		Cronbach's alpha		.789

5.2.2 Interfaces: Conversational agent (CA) and remote control unit (RCU)

The CA penetration rate per household in South Korea is approximately 11% as of April 2018 (Consumer Insight, 2018). According to the report, the most widely distributed CA service is Giga Ginnie (39%), which is combined with the Internet Protocol TV (IPTV) set-top box and serves as an assistant in watching TV (Figure 5.2.1).

Figure 5.2.1 TV remote control (left) and conversational agent system (right) for Giga Ginnie.

The method to use of Giga Ginnie consists of wakeup (e.g., “Hey Ginnie”) and response (“Yes”) → command (e.g., “What time is it?”) → execution/response sequences, similar to that of other CAs (see Fig 5.2.3). In such a method, as Reeves (2017) had mentioned, speaker selection issue or question-answer pairs issue exists. For example, if the wakeup and command sequences are exchanged, (e.g., “What time is it now, Ginnie?”), the command spoken before the wakeup command is ignored (speaker selection issue). Furthermore, when a speaker says commands consecutively or provides a feedback for a response of the CA, there is a constraint that the commands are recognized only if the wakeup (“Hey Ginnie”) and response (“Yes”) steps are performed (question-answer pairs issue).

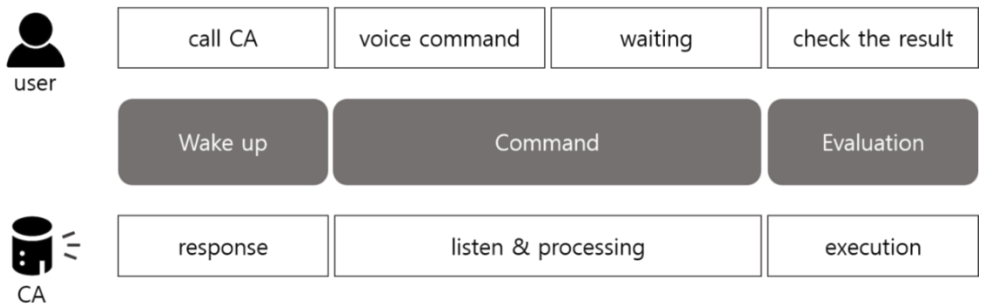


Figure 5.2.2 The interaction process between user and conversational agent.

In the case of the rate of semantic interpretation for voice recognition, the recognition rate of simple commands composed of objects and verbs only (e.g., “Turn on the TV,” or “Show today’s weather”) was more than 90% on average. However, when a command contained two or more actions (e.g., “Turn on the TV and show today’s weather”), the result could not be predicted. Either CA performed all commands or performed only one. Considering these problems, tasks were created based on objects and verbs so that they can be executed properly. When the execution step is complex, the task is divided into subtasks. The difference in voice recognition rate among users was not large but exists. The recognition rate was slightly different depending on the voice tone, height and volume. In order to reduce the gap, I made a voice guide for interaction with CA and provided it to the participants to practice before the experiments.

5.2.3 Physiological measurements

Many studies asserted that the user “emotion” plays an important role in the UX of entertainment products; however, it is very difficult to measure emotions and quantify them. Emotion data are usually obtained from results of subjective responses (self-report questionnaires), thinking aloud, heuristic evaluation, and

contextual inquiry. Because it relies on the user's recall or subjective expression, which makes it difficult to measure objectively and focuses on the result rather than the process (Wilson & Sasse, 2000b). The emotional assessment should also consider a quantitative data measurement method, such as usability testing, which uses data based on an objective standard such as performance time or error frequency. One of the methods proposed to address these issues is physiological measurement (Kreibig, 2010; Mandryk, Inkpen & Calvert, 2006; Yao et al., 2014).

Physiological measurement is a method generally used to determine human emotions in the psychology area, and includes heart rate (HR), skin conductance (SC), and facial muscle activity measurements. It involves the measurement of cognitive effort or emotional response by attaching a sensor on the face or the body and sensing the physiological state (frowning, smiling, perspiration, etc.) signals appearing as a result of interaction (Ekman, 1999). Levenson (1992) and Hamdi et al. (2012) asserted that physiological data sources clearly produce different results from each other depending on the emotions. In a physiological measurement experiment targeting game users, Mandryk, Inkpen and Calvert (2006) presented a result that the physical responses between games having a computer as opponent and games having a friend as opponent were different. They determined that a more specific and an accurate evaluation would be possible through a cross-verification of quantitative data and subjective data obtained directly from users. Yao et al. (2014) also collected physiological data and self-report data together in a UX evaluation of a mobile application. They asserted that through a result showing correlations between SC results and self-report data, physiological measurement can be one of the effective methods for UX evaluation.

Emotions (anger, confusion, pleasure, etc.) felt by the experiment participants while

interacting with the TV interface (CA and RCU) are shown by changes in body signals (e.g., smiling, eye frowning, and increased HR). Physiological experiment is a method of collecting and analyzing signals that appear during such body changes. These results make sense that the results provide objective and quantitative data and has the advantage of being able to see specifically how the participants' emotions changes occur at the moment of interaction. This study collected the data through electrocardiography (EKG), electromyography (EMG), and skin conductance (SC) measurements (Fig 5.2.4). The EKG is known to increase due to not only positive emotions (e.g., happiness and pleasure) and surprise, but also negative emotions (e.g., anger, anxiety, embarrassment, fear, and sadness) (Jerritta, Murugappan, Nagarajan, et al., 2011). The collected EKG data were converted to HR data again and analyzed. The EMG measures the movement of the facial muscles when the user receives emotional stimulation by attaching sensors between the eyebrows or under the eyes.

According to the related studies, the EMG data are generally related to negative emotions (Nakasone, Prendinger, & Ishizuka, 2005). When people respond to psychological stimulations, the SC measures the changes in skin conductance, particularly signals of certain sweat glands on the finger or the foot. The SC data result is regarded as an index of arousal, anxiety, or stress (Hudlicka, 2011). Yao et al. (2014) also insisted that SC data were higher in the failed tasks than the successful tasks. The reason for adopting the physiological measurements is mainly to collect the objective data by measuring the user's body reaction signals depending on the emotions. Another important reason is that it is possible to grasp specific emotional changes during the interaction process. Most emotional experience results were gathered after experiments. However, it was difficult to know precisely about the emotional changes and the points during the interaction. In particular, since interacting with CA is based on the series of conversations, it is important to

understand the emotional changes that users feel over time, from the moment of utterance to the response. The motions of participants and the TV screen that responds to these were all recorded in videos and used as reference data in the analysis process. Physiological data were collected and analyzed with AcqKnowledge 4.1 software from Biopac Systems, Inc.

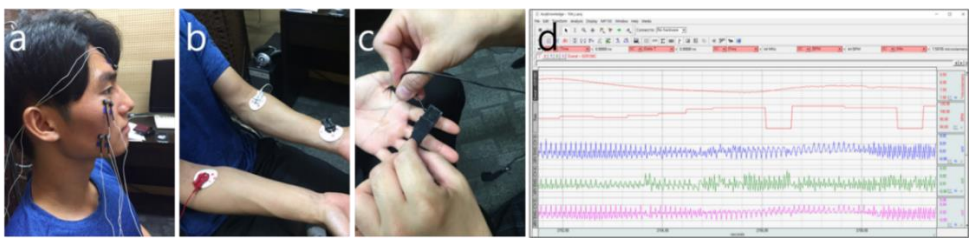


Figure 5.2.3 Sensors for gathering physiological data. (a) EMG; (b) HR; (c) SC; (d) AcqKnowledge 4.1 software.

5.2.4 Ergonomic quality and hedonic quality measurements for UX

According to Hassenzahl, Diefenbach & Göritz (2010), a product's character is a set of product attributes that are created by a combination of product features and user expectation or standard. A product's character can be explained mainly by the following two attributes: pragmatic attributes (e.g., efficiency) and hedonic attributes (e.g., pleasure). Pragmatic attributes include usability factors such as usefulness and effectiveness, and hedonic (= emotional) attributes indicate emotional factors such as fun and boredom. Hassenzahl and his colleagues (2010) proposed the AttrakDiff2 questionnaires to measure the pragmatic attributes and hedonic attributes. AttrakDiff2 consists of four quality dimensions, namely pragmatic quality (PQ), hedonic quality (HQ)-satisfaction (HQ-S), hedonic quality-identification (HQ-I),

and APPEAL. It indicates the overall experience satisfaction and has a method of choosing a level similar to a person's own mood state by having mutually opposing (semantic differential) adjectives on both sides. After finishing each experiment, the participants evaluate subjectively the AttrakDiff2 questions that ask about their own feelings. The PQ consists of seven items for the usefulness aspect such as "complicated–simple" and "unpredictable–predictable." The HQ consists of seven HQ-S items and seven HQ-I items. The HQ-S consists of "conventional–inventive," "dull–captivating," etc., and the HQ-I consists of emotional adjectives such as "isolating–connective" and "tacky–stylish." ATT refers to the overall UX, which combines the perceived HQ and PQ, and consists of seven items such as "repelling–appealing" and "bad–good." Each item consists of seven scales from −3 to +3, and a higher value indicates that the user has a more positive experience.

5.3 Results

Among the physiological response results, the cases in which videos were difficult to replay due to controller mistake were excluded, and the measurement data of 37 persons in the CA session and data of 35 persons in the RCU session were analyzed. As for the measurement target tasks, 9 tasks were chosen as analysis targets by excluding the tasks in which the performance time was difficult to calculate (T1, T2, T3, and T12), and which showed an extremely large time difference between the CA and RCU sessions (T5). In the RCU session, two cases were not recorded due to video recorder controlling error. For the task measurement scope, the time from the utterance (or clicking a remote controller button) to screen switching time point was estimated to be 5 seconds, and based on the screen switching time point (7 seconds afterward and 13 seconds beforehand), total 20 seconds were set as the task scope.

The analysis was performed for this section. For Task 7, which has complex performance steps, a 60 s section of 7 s after and 53 s before the screen switching was analyzed. Among the measurement target tasks, the success cases were analyzed first, and the cases that failed were categorized separately and comparatively analyzed with the physical response changes of the experiment participants to determine if there were differences between the interface controls. For the EMG data among the physiological measurement records, the originally collected data were downgraded with 50 sample frames per second, and afterward, 1,000 signals were extracted from a section corresponding to 20 s. Then, analyses were carried out based on these data. Fig 5.3.1 shows the overall results of each task completion rate. The reason for the failure of the tasks were mainly the participant failed to input the text or could not find the menu in the RCU session. In CA sessions, most of the task completion failures were due to the voice command recognition errors. Between the RCUs and CAs, the task completion rate was higher in the RCU session (RCU = 96.6%; CA = 93.75%).

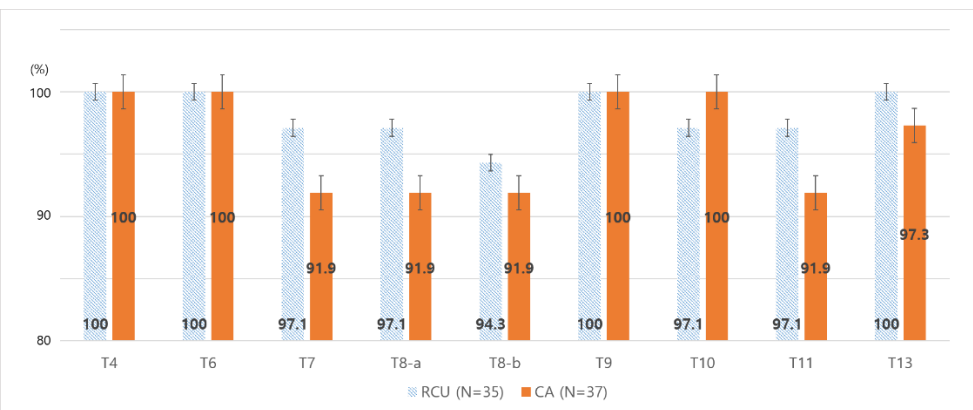


Figure 5.3.1 Task completion rates of RCU and CA

5.3.1 Physiological measurement results

For each task performed in the CA and RCU sessions, repeated measurement analyses were carried out to examine if there were differences in SC measurement values for each task performed. For all the tasks excluding Task 7, the data of 13 s were extracted from the data that recorded 20 s by excluding the dummy data at the front part, and the difference between the interfaces was measured. The SC result showed significant differences partially or overall in the 7 tasks among the total of 9 tasks (see Table 5.3.1).

Table 5.3.1 SC and HR results in the experiments with CA and RCU.

Tasks	Model	SC			HR		
		Mean Sq.	F	Sig.	Mean Sq.	F	sig.
T4	Interface	.055	.072	.790	2327.540	.692	.411
	Time	.023	.719	.733	220.058	.566	.869
	Interface*Time	.047	1.079	.376	1334.960	2.640	.002**
T6	Interface	2.719	4.781	.036	46.027	.012	.912
	Time	.077	3.408	.000***	278.396	.689	.763
	Interface*Time	.144	7.080	.000***	276.874	.803	.648
T7	Interface	14.810	1.471	.234	5662.037	1.036	.317
	Time	.112	1.562	.007**	255.085	.736	.920
	Interface*Time	.066	1.203	.155	344.467	.948	.581
T8a	Interface	.274	.427	.519	291.560	.248	.622
	Time	.063	1.644	.078	148.131	.515	.905
	Interface*Time	.063	3.049	.000***	359.552	.940	.506
T8b	Interface	.248	.304	.586	7404.521	3.382	.076
	Time	.060	1.967	.026*	370.552	.970	.477
	Interface*Time	.033	1.669	.072	347.664	.932	.515
T9	Interface	.694	.902	.349	1722.355	.812	.374

	Time	.116	3.946	.000***	371.191	.898	.549
	Interface*Time	.011	.554	.878	547.172	1.301	.215
	Interface	.317	.356	.555	3.628	.001	.972
T10	Time	.037	.781	.671	153.991	.370	.973
	Interface*Time	.302	5.245	.000***	800.609	1.919	.031*
	Interface	.263	.474	.497	2978.173	.775	.386
T11	Time	.040	1.602	.089	350.442	.978	.469
	Interface*Time	.018	.698	.754	294.377	.767	.684
	Interface	8.939	5.131	.03*	1.217	.000	.983
T13	Time	.089	2.033	.021*	209.757	.698	.754
	Interface*Time	.158	4.009	.000***	476.760	1.698	.065

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

SC changed significantly over time in the T6, T7, T8b, T9, and T13 tasks of the CA session. T6, T8a, T10, and T13 were the tasks that showed differences by interface and time. The most common graph shape for CA was a high SC score at the point of utterance which then decreased over time (see Figure 5.3.2a). Figure 5.3.2 shows the changes in SC. Overall, it indicates a significant difference in SC between RCU and CA. CA showed the highest SC value at the start and gradually decreased, while RCU showed no significant change or rather increased over time. The results imply that participants' emotions responded differently even though they performed similar tasks. For example, Task 6 concerns content retrieval. In the RCU session, it was necessary to click the RCU button repeatedly to find the content title, which is at a lower level. Task 8a also took time because it required going through several stages to reach the goal. In T10 and T13, the SC increased over time in the RCU session. In Task 10, the content's title was hidden so that the participants might feel difficulty in predicting the destination. Task 13 was a very easy task with the TV off, but the feedback points were different. In the CA session, after the participant gave

the command to turn off the TV, the CA's "Turn off the TV" feedback appeared first and then the TV screen blacked out. In the RCU session, however, after the RCU's power-off button was pressed, the TV screen blacked out without any additional feedback. On the other hand, in the CA session, it was assumed that cognitive effort is required at the initial stage of interaction. In all tasks, the SC was highest at the beginning of the interaction and then decreased over time. A previous study (Nakasone, Prendinger, & Ishizuka, 2005) indicates that high SC values are associated with negative emotions. Therefore, it is speculated that cognitive effort by the user is required at the voice-command stage and that negative emotions are caused by this. The reason for the decrease in SC is that the tension is relieved and the uncertainty disappears when feedback is received immediately after the voice command is given.

In Figure 5.3.3 below, the SC measurement results for each task are graphed. Apart from the significant differences between the two interfaces, the SC data for each task showed different results for the two interfaces. Task 6 showed significant differences in both time and interface with menu navigation. The CA goes through the menu search by voice, whereas the RCU navigates the menu via the click of a button. In this situation, it is predicted to have a cognitive load in the process of thinking commands in the CA interaction. The graph in RCU interaction implies that repeated button clicks cause negative emotions during the process. Task 8a is regarding VOD control. The result of RCU interaction shows no significant change in the graph, but the result of CA shows that the initial stage of CA interaction is highly tense. Similar to Task 6, Task 11 is a menu navigation task. Compared to

CA, RCU does not change significantly in this session. In contrast, CA has a very high common initial stress.

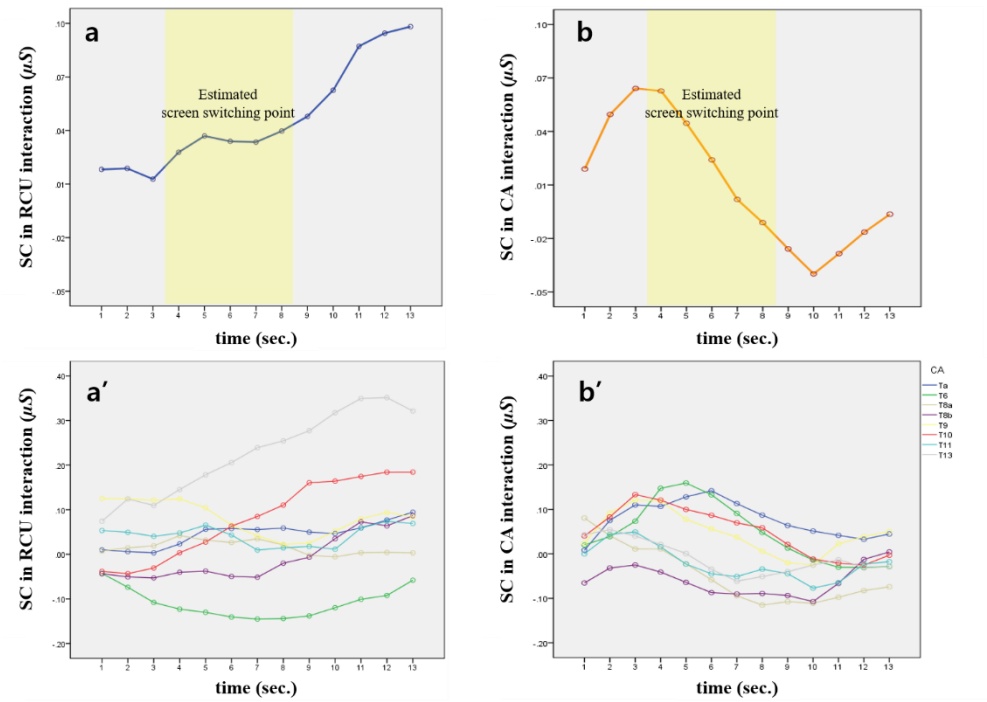


Figure 5.3.2 Upper side shows interaction between SC and RCU (a) and SC and CA (b). SC data for each task in RCU interaction (a') and CA interaction (b') are shown below.

The graph in Task 10 shows the most distinctive contrast between the two interfaces. In the case of CA, it is necessary to envision how a content recommendation command is to be issued. In contrast, in the case of the RCU, the user needs to click a button several times to enter a search term.

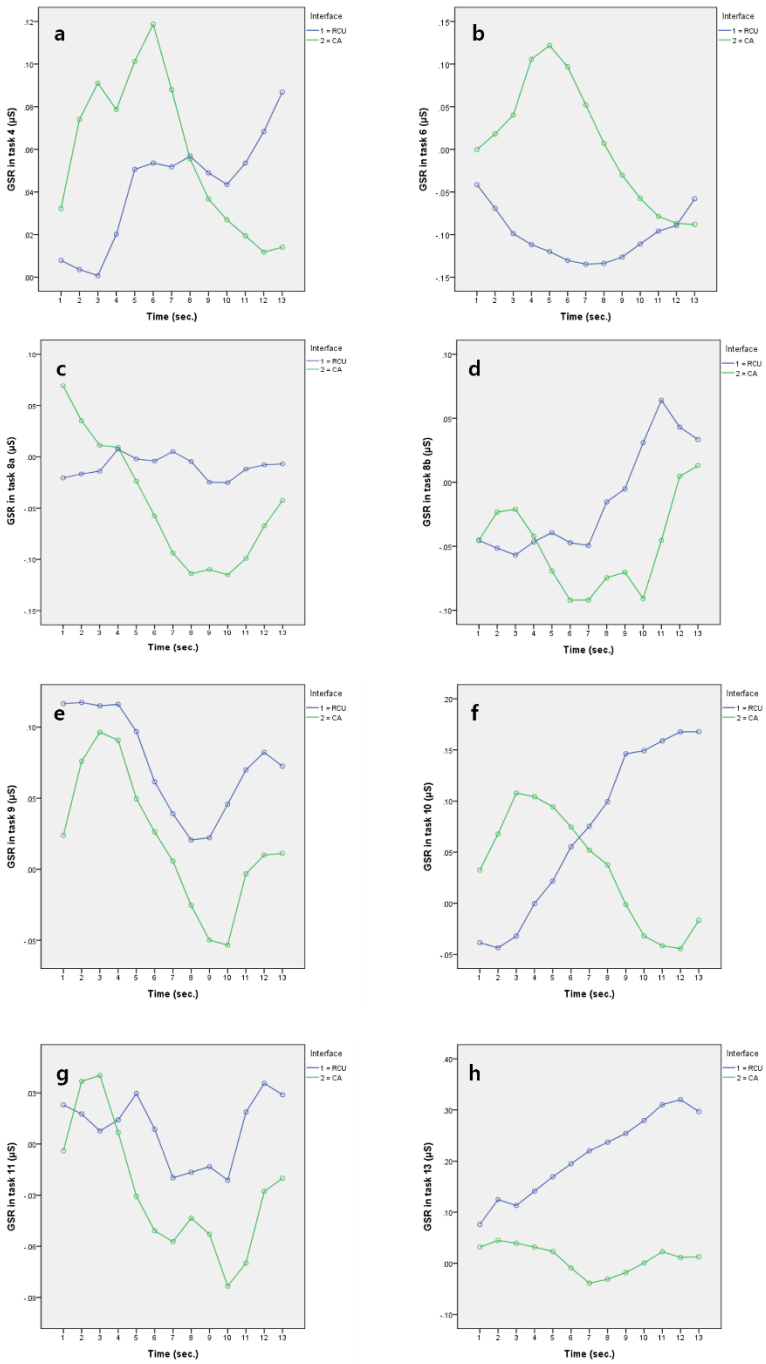


Figure 5.3.3 SC graphs for each task: (a) T4, (b) T6, (c) T8a, (d) T8b, (e) T9, (f) T10, (g) T11, and (f) T13.

Table 5.3.2 The results of SC data analysis by tasks

Category	Task		Graph	
			RCU	CA
1. Low to high in RCU	Channel zapping	T4	↗ (low → middle)	
	Menu navigation	T10	↗ (low → high)	
	VOD control.	T8b.	↗ (low → high)	
	TV off	T13	↗ (low → high)	
High to low stress in CA	Channel zapping	T4		↘ (high → low)
	Search	T6		↘ (high → low)
	Search	T10		↘ (high → low)
	Menu navigation	T11		↘ ↗ (high → low → middle)
	VOD control	T8a.		↘ ↗ (high → low → middle)
	VOD control	T8b.		↘ ↗ (middle → low → middle)
Same	Channel zapping (repeated task)	T9	↘ ↗ (middle → low → middle)	↘ ↗ (middle → low → middle)
No change	Menu navigation	T6	→(No change)	
	Menu navigation	T11	→(No change)	
	VOD control	T8a.	→(No change)	
	TV off	T13		→(No change)

After comparing the results between RCU and CA, commonalities for each graph type were also analyzed (see Table 5.3.2). Most of the results of CA interactions show high initial stress and then low, regardless of the level of difficulty. RCU showed no change in stress (T6, T8a, T11) or initially in a low level and increased stress over time (T4, T8b, T10, T13). In T10, the RCU and CA graphs showed completely opposite results. T10 is about searching the TV content. In RCU interaction, the user has to click the menu several times to search the TV content. On the contrary, in CA interaction, the process was very short but required cognitive efforts in the process of conducting the task. T9 is a task for channel zapping. The graphs of two interfaces showed almost the same results: decreasing somewhat and then rising slightly. This implies the users performed the task without stress because they were accustomed to that which was conducted before.

The implications of the recorded SC results are as follows: In CA interactions, cognitive effort is necessary at the beginning for the process of constructing complex commands. This seems to be due to the fact that most of the participants were not

familiar with controlling a TV with CA, and they may have been burdened by following the voice command sequences. Clicking a button on a remote control, on the other hand, although repetitive, has not been regarded as mentally burdensome. However, the graphical results indicate that as the process of reaching the destination becomes longer, negative feelings such as uncertainty or anxiety increase. The HR data showed statistically significant differences in Task 4 (interface * time: $F(12, 408) = 2.640$, $p < .01$) and task 10 (interface * time: $F(12, 396) = 1.919$, $p < .05$). The characteristics of heart rates data, however, showed no significant change over time as in SC, but showed a contrast between the two interactions according to time. No significant difference was found in the EMG data.

5.3.2 UX evaluation results

Repeated measurement analyses were conducted to compare if there are differences between the interfaces (CA and RCU) in terms of UX dimensions (PQ, HQ-I, HQ-S, and ATT). Consequently, the PQ showed a significant difference, and a correlation was shown between the interface and UX dimensions. Furthermore, significant differences were shown between the HQ-S dimension and the interfaces in terms of interface, UX dimensions, and interface × UX dimensions. In the HQ-I dimension, there were differences between the interfaces, and a correlation was shown as well. In the ATT, there were significant differences between the two interfaces, but no correlation appeared (see Table 5.3.3).

Table 5.3.3 Repeated measures ANOVA table reporting F-values for UX dimensions according to interface (RCU and CA).

Dimension	Model	Mean square	F	sig.
PQ	Interface (1 = RCU, 2 = CA)	36.125	3.072	.087
	Interface*PQ	20.658	14.026	.000***
HQ-I	Interface	74.650	10.808	.002**
	Interface*HQ-I	2.251	2.536	.021*
HQ-S	Interface	440.782	70.363	.000***
	Interface*HQ-S	13.477	14.295	.000***
APPEAL	Interface	51.540	7.147	.011**
	Interface*ATT	1.390	2.514	.022*

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

A two-tailed paired t-test was conducted to compare the difference between the CA interaction and the RCU interaction for significance (see Table 5.3.4). Among the perceived pragmatic quality dimension, “predictable” ($t(40) = 4.394$, $p < .001$), “clearly structured” ($t(40) = 3.483$, $p < .05$), and “manageable” ($t(40) = 4.833$, $p < .001$) were significantly larger than in the CA interaction. The mean value of “human” ($t(40) = -.2816$, $p < .05$), one of pragmatic quality attributes, was significantly higher in the CA interaction than in the RCU interaction. In summary, participants responded that CA felt more like a human than RCU. At the same time, they had a much more unpredictable, confusing and unruly experience in the CA interaction.

Table 5.3.4 Paired t-test results for RCU and CA.

		RCU mean (SD)	CA mean (SD)	t	sig. (2-tailed)
PQ	technical - human	-1.29 (1.38)	-.27 (1.49)	-2.816	.008**
	complicated - simple	.02 (1.81)	.37 (1.46)	-.883	.382
	impractical - practical	.15 (1.84)	.07 (1.65)	.169	.867
	cumbersome - straight forward	-.02 (1.81)	-.46 (1.53)	1.037	.306
	unpredictable - predictable	1.27 (1.50)	-.24 (1.41)	4.394	.000**
	confusing - clearly structured	.90 (1.56)	-.32 (1.05)	3.483	.001**
	unruly - manageable	1.51 (1.42)	-.12 (1.37)	4.833	.000**
HQ-I	isolating - connective	-.07 (1.44)	.54 (1.30)	-1.863	.070
	unprofessional - professional	-.39 (1.32)	-.24 (1.20)	-.567	.574
	tacky - stylish	-.34 (0.99)	.56 (1.30)	-3.287	.002**
	cheap - premium	-.54 (1.12)	.66 (1.12)	-5.267	.000***
	alienating - integrating	-.24 (0.97)	.49 (1.16)	-2.877	.006**
	separates me - bring me closer	-.12 (1.25)	.78 (1.17)	-3.029	.004**
	unpresentable - presentable	0.00 (1.50)	.56 (1.57)	-1.485	.145
HQ-S	conventional - inventive	-1.80 (0.95)	1.29 (1.19	-11.667	.000***
	unimaginative - innovative	-1.27 (1.04)	.54 (1.45)	-5.503	.000***
	cautious - bold	-.17 (1.28)	.17 (1.24)	-1.180	.245
	Conservative - innovative	-1.10 (1.09)	.78 (1.15)	-5.826	.000***
	dull - captivating	-.98 (1.00)	.59 (1.18)	-5.160	.000***
	undemanding - challenging	-.73 (1.30)	.85 (1.09)	-5.962	.000***
	ordinary - novel	-1.71 (0.85)	.29 (1.05)	-8.634	.000***
ATT	unpleasant - pleasant	-.80 (1.04)	.24 (1.47)	-3.066	.004**
	ugly - attractive	-.05 (0.74)	.63 (1.22)	-3.285	.002**
	disagreeable - likeable	.02 (0.80)	.68 (1.14)	-2.402	.021*
	rejectable - inviting	.22 (1.10)	.93 (1.42)	-2.220	.032*
	bad - good	.39 (0.81)	.76 (1.10)	-1.416	.165
	repelling - appealing	.10 (0.67)	.37 (1.02)	-1.171	.248
	discouraging - motivating	-.20 (0.89)	.27 (1.13)	-1.934	.060

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

The result of the hedonic quality dimension was different from that of pragmatic quality dimension. There were more differences in hedonic quality dimensions than in pragmatic quality, especially in the hedonic quality stimulation dimension. Among of hedonic quality identification dimension, “stylish” ($t(40) = -3.287$, $p < .001$), “premium” ($t(40) = -5.267$, $p < .001$), “integrating” ($t(40) = -1.877$, $p < .05$), and “brings me closer” ($t(40) = -3.029$, $p < .05$) attributes showed significant differences. Although not significant, other attributes were also higher in CA than in RCU. In the hedonic quality stimulation dimension, the scores of CA were also higher than those of RCU for all attributes. Except for the “bold” attribute, all other attributes showed significant differences, and the mean differences between the two interfaces were larger than those of the identification dimension.

Figure 5.3.4 shows the visual differences between the two interaction experiences. The results of pragmatic quality evaluation at the top of the graph shows a much more positive experience with the RCU. During interacting with RCU, only “technical – human” attribute was negatively experienced. The middle of the graph shows the hedonic quality evaluation results. According to the results, interacting with RCU were evaluated negatively in all attributes, while CA was positively experienced in almost all attributes. APPEAL is a combination of PQ and HQ, indicating the degree of overall user experience. In the APPEAL dimension, the mean values of CA were higher than those of RCU for every aspects. “Pleasant” ($t(40) = -3.066$, $p < .01$), “attractive” ($t(40) = -3.285$, $p < .01$), “likeable” ($t(40) = -2.402$, $p < .05$) and “inviting” ($t(40) = -2.220$, $p < .05$) attributes of the CA showed significant differences to those of RCU. To sum up, the participants responded that although the CA was unpredictable, confusing, and difficult to control, it was, at the same time, more connected and closed (HQ-I dimension), and new and creative (HQ-S dimension). In the APPEAL dimension, which indicates the overall

user experience, participants answered they had more pleasant, attractive, and enjoyable experiences with CA than RCU.

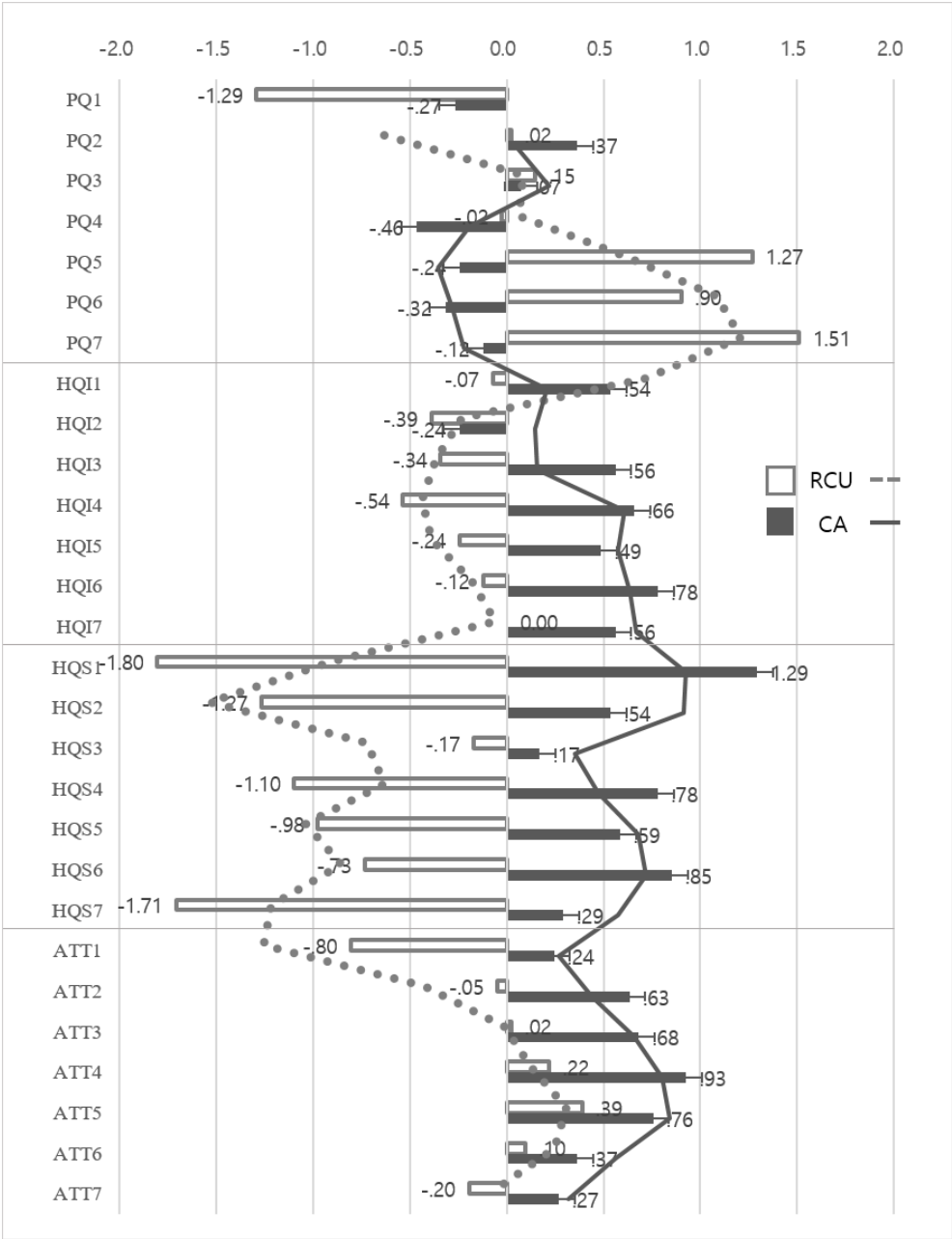


Figure 5.3.4 Mean differences of UX attributes between RCU and CA.

5.3.3 The correlations between UX attributes and APPEAL

A linear regression was calculated to predict APPEAL of each interface based on UX sub dimensions (see Table 5.3.5). The result of Durbin-Watson test was 2.448, which was close to 2, so it was judged to be suitable for multiple regression model. In the RCU session, a significant regression equation was found (F (3, 37) = 22.814, $p < .001$), with an R^2 of .649. As a result of confirming the significance of each dimension, it was confirmed that PQ (t(40) = 5.313, p < .001) and HQ-S (t(40) = 2.072, p < .05) excluding HQ-I had an effect on APPEAL. Both PQ (B = .401) and HQ-S (B = .268) were positive and the influence of PQ (β = .597) on APPEAL was higher than HQ-S (β = .255). The explanatory power explaining the dependent variable by the independent variables was 64.9%. In the CA session, a significant regression equation was found (F (3, 37) = 16.387, $p < .001$), with an R^2 of .571. As a result of confirming the significance of each dimension, only HQ-S (t(40) = 2.111, p < .05) affected on APPEAL. The explanatory power explaining the dependent variable by the independent variables was 57.1%.

Table 5.3.5 Results of linear regression analysis for APPEAL of RCU and CA based on UX attributes.

		R	R^2	R^2(adj)	B	S.E	β	t	sig.
APPEAL (RCU)	PQ				.401	.076	.597	5.313	.000***
	HQ-I				.264	.131	.264	2.006	.052
	HQ-S				.268	.129	.255	2.072	.045*
		.806	.649	.621					
APPEAL (CA)	PQ				.297	.153	.303	1.945	.059
	HQ-I				.225	.176	.226	1.277	.210
	HQ-S				.365	.173	.327	2.111	.042*
		.755	.571	.536					

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

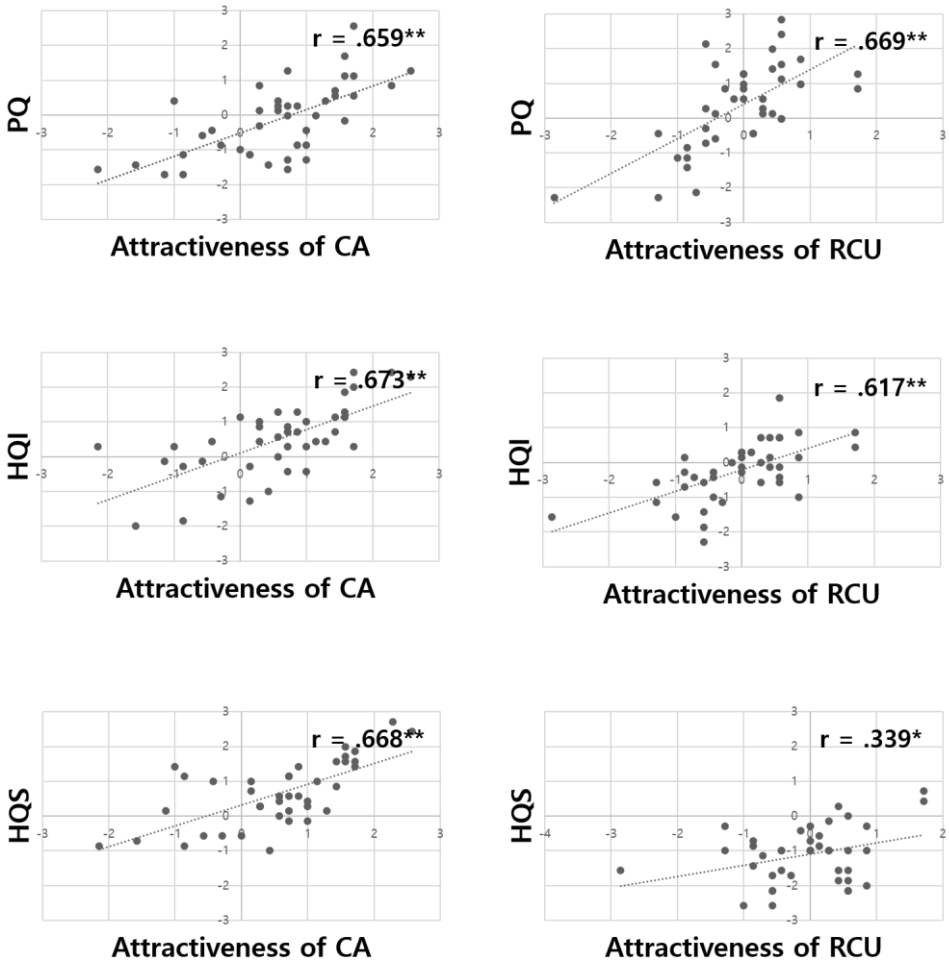


Figure 5.3.5 Correlations of APPEAL and UX quality dimensions. The upper part shows the correlations between ATT and UX sub-dimensions of CA (from the left, ATT x PQ, ATT x HQ-I, and ATT x HQ-S). The lower part shows the correlations between ATT and UX sub-dimensions of RCU (from the left, CA (from the left, ATT x PQ, ATT x HQ-I, and ATT x HQ-S).

Similarly, the correlation analysis results also showed a positive linear relationship between APPEAL and other UX quality dimensions (PQ, HQ-I, and HQ-S). In

Figure 5.3.5, the upper side shows the correlation analysis results between ATT and UX sub-dimensions of CA (ATT x PQ: r = .659**, ATT x HQ-I: r = .673**, and ATT x HQ-S: r = .668**). The lower side shows the correlation analysis results between ATT and UX sub-dimensions of RCU (ATT x PQ: r = .669**, ATT x HQ-I: r = .617**, and ATT x HQ-S: r = 339*).

5.3.4 The correlations between SC and UX attributes

In the correlation analysis between SC data and each UX dimension, there was a significant negative correlation between SC and PQ in RCU interaction (r = -.314*, $p < .05$). No significant correlation was found in CA interaction. In order to investigate the correlation between SC and UX attributes in detail, correlation analysis was conducted between SC and each UX attribute. Table 5.3.6 shows the correlation analysis results. A total of 33 correlations were obtained from each attribute and SC data correlation analysis. However, only the results more than two times were analyzed because it was difficult to find consistency in one time occurred correlation. In the RCU session, a negative correlation was found between the SC data and “straight forward” attribute (T8a, T9, and T11). Positive correlation with 'Stylish' appeared in T6 and T9. In the CA session, the "human" attribute in T8b, T9, T10, T11, T13 and "connective" attribute in T8a, T8b, T9, T10, T11 showed common positive correlations. The "appealing" attributes showed negative correlations in T8a, T8b, T9, T10, T11, and T13. Correlation between physiological data and UX measurement results was greater in CA interaction than in RCU interaction. In summary, properties that negatively affect SC data in RCU use are considered cumbersome. In CA interaction, the feeling of being human or connected is considered positive, and repelling is considered as a negative emotional attribute.

Table 5.3.6 Correlations between SC and UX attributes in RCU and CA interaction, respectively.

	UX attributes	SC								
		T4	T6	T7	T8a	T8b	T9	T10	T11	T13
RCU	cumbersome - straight forward				**-.369***		**-.400****		**-.357***	
	tacky - stylish		.337*				.323*			
CA	technical - human					.302	.314*	.350*	.366*	.327*
	isolating - connective				.319*	.315*	.312*	.335*	.325*	
	repelling - appealing				**-.438****	**-.391***	**-.382***	**-.334***	**-.345***	**-.344***

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

5.4 Conclusion

In this study, I explored how a conversational agent affects a TV viewer's user experience by comparing CA and RCU. Participants performed tasks related to TV watching by interacting with a CA and by interacting with an RCU. After the experiments, they self-assessed the experiences that they had during the interactions. In addition to this, physiological indicators were used to measure emotional responses by collecting the signals corresponding to the participants' physical reactions during the interactions. This study yielded some valuable results, but at the same time, they were influenced by several unexpected factors in the experiments. I conducted a pre-check on the CA performance in advance and configured the tasks to avoid errors. Nevertheless, wrong results were provided in some cases because the voice commands of users were not recognized or were incorrectly understood (machine response error) in the CA interaction. Despite the fact that the "wakeup → answer → command" sequence was sufficiently explained to the experiment

participants through guidance beforehand, cases of saying wakeup and command consecutively were observed (user input error). This seems to be a result of showing ordinary conversation habit unconsciously.

Occasionally, a case of expressing a wrong result was observed because a different command was spoken instead of the appropriate command for the task performance. In Task 7, which had the highest difficulty level, there were more cases of succeeding a task with two or three attempts than the cases of succeeding by saying a command only once. Furthermore, in some cases, even when a content was successfully searched (object–verb relationship), failure occurred in the process of moving from the search result screen to the next stage. The execution commands such as “confirm” and “play” were not used much verbally because they are the functions commonly used with the RCU. In the CA session of this experiment, only the voice was used without the RCU to complete the tasks. However, these execution commands had lower recognition rates than the other words and errors of showing wrong results occurred more. One participant tried to perform a task for more than 10 min and gave up. It seemed that the reason why the recognition rate of execution commands such as “confirm” was low was because they were not commonly used much and the learning was not sufficient. The task performance failure rate of the CA session was higher than that of the RCU session, and the confusion experienced while performing these tasks was reflected in the results of questionnaires that were evaluated by the participants themselves after the experiment. In the ergonomic quality dimension, the scores of CA were significantly lower than those of RCU for the “difficult to predict,” “confusing,” and “difficult to control” factors.

The ANOVA analysis of the subjective evaluation results showed that the UX qualities vary according to the interaction (CA and RCU). The paired t-test, which

compares the quality of each US attribute, statistically confirmed the difference between the CA and the RCU. CA was inferior to RCU in all attributes except "human" in pragmatic quality dimension. Hedonic quality was higher in CA than in RCU in both HQ-I and HQ-S. APPEAL, PQ and HQ combined UX dimension, was also higher in CA than in RCU. Regression analysis between APPEAL and other quality dimensions revealed that PQ was the most influential factor in RCU session. HQ-S mostly influenced on APPEAL in CA session. In sum, the participants responded that they felt that the interactions with the CA were more interesting and attractive although they felt frustrated and confused because the goal was not accomplished or the conversation was not properly carried out.

The SC data suggested that arousal level has significantly increased when the number of button clicks increased, or when the result of the task to be executed was not clearly predictable. During the experiment design stage, I regarded CA as an alternative to RCU and compared users' interactions between CA and RCU. However, what I found through this experiment is that CA and RCU can complement each other. From the SC data, it can be seen that in the RCU session, the SC data is maintained at a constant level or even increases over time. This suggests that users get aroused when they are pushing the remote controller's buttons repeatedly, which has likely caused frustration in them. It also suggests that being unable to predict the outcome before reaching the final result also leads to arousal, possibly caused by anxiety.

On the other hand, in the CA session, the SC value during the initial stage of the instruction was stable, and then the result was significantly different from the RCU session result. The high level of SC suggests the increased stress level of participants caused by the heavy burden of the task; having to follow a complicated command

protocol. The rapid lowering of SC suggests the relief of the tension as feedback via television screen is provided. The important point is that the areas where CA and RCU have advantages are different, and the emotions that users feel in these interaction processes are also opposed to each other. These results can also affect TV operation scenarios and TV function implementations. For example, in terms of functionality, CA is better suited for quick-access tasks of jumping directly to a destination rather than multiple button clicks (e.g., text input) or semantic search (e.g., “find a comedy movie”). The RCU is better suited for performing tasks that are dense in a short range, such as meticulous manipulation (e.g., “fast forward” or “rewind”), configuration, or menu navigation. Through these results, it was discovered that the experience with the CA was more positive than that with the RCU in the emotional dimension. In addition, it was confirmed that fun and calm factors had effects on the positive experience of TV viewers. This was shown in not only the subjective evaluation results, but also the physiological measurement results. The fact that the approach of interacting with the voice is more fun and more emotionally pleasing is considered to be an essential part of the application experience. Currently, the CA function has been added without changing the existing TV viewing context. However, in the future, it is expected that the interface design and the scenario for utilizing the functional aspect of the CA and its emotional properties will change. It is expected that the TV viewing context will be changed and that the TV experience will be strengthened.

There is a partial correlation between the results of the physiological measurements and the UX quality evaluation results. In the CA sessions, “human” and “connective” attributes of UX showed a positive correlation with SC data, and “appealing” showed a negative correlation with SC. A negative correlation between “straight forward” and SC data was showed in the RCU sessions. I measured SC, EMG, and HR as

biometric markers, but only SC gave meaningful results. HR data also showed a partial difference between CA and RCU, but it was less applicable than SC, and no clear interpretation was found. In the EMG measurement results, no significant difference was also shown between the CA and the RCU. This result can be interpreted in several ways. First, after attaching sensors at the eyebrow, around the eye, and around the mouth on the face, the EMG measures changes in facial muscles that are altered depending on the feeling of the user. However, these changes are difficult to measure accurately because the changes in mood appear diversely from a case that appears right away to a case that shows a slight change or no change at all. Furthermore, in the CA session, because the mouth has to be moved to say a command, the difference between the section of change according to feeling and the command section cannot be distinguished or an interference with each other can occur. Considering this point, EMG signal measurement method should be carefully considered in the emotional measurement through the conversational agent, or a precise analysis is needed to distinguish the signal interference.

The voice recognition function will be combined indispensably with various products around us in the near future. This is because it is not just “smart” but also a method that “stimulates emotion” and more natural interaction. In the HCI area, the usability perspective that has been accumulated for a long time has extended to the UX framework as the importance of emotional quality has risen. On top of this, the active adoption of natural interactions of humans such as hearing and tactile sense, which go beyond the visual interaction, requires UX studies in more multidisciplinary and in-depth perspective. In this aspect, it will be a meaningful attempt to use the physiological measurements to directly determine the physical responses occurring immediately in the interaction process. People think that they know their emotions or feelings and express them properly, but in reality, it is not

true. Particularly, it is almost impossible to remember completely or determine accurately the instantaneous changes of emotion or the feelings that change every moment. Natural interaction requires a physical response measurement, and through this, full explanation can be provided on how the UXs were constructed. Nevertheless, for the part regarding which result will be obtained by applying which method, diverse studies are required according to the target situation.

The technological evolution of the CA is ongoing and at the same time, various perspectives are being considered. In the 2016 IBM Developer Day, Michal Yuan (2016) emphasized two things for success factors when designing a CA system: the CA has to be operated in a limited scope, and by providing characteristics suitable for the usage scope and goal of the CA, a context-model design for conversation with the user should be properly done. Other parts that should be considered along with the technological evolution of the CA are non-linguistic expressions such as sighs, hand gestures, and facial expressions. These are major signals that are referenced when I communicate with a counterpart. In order to infuse the CA naturally into the daily lives of users, studies on various non-linguistic expressions (e.g., tones and nuances) should be carried out in parallel with efforts to improve the recognition rate of voice. If personified cues (anthropomorphic or personification cue) are added, more natural and ordinary interactions will be possible. Through such a process, the voice recognition technology will be combined with our everyday products as an "intelligent agent" instead of just being "smart" interaction tools. Although a review on the effective implementation method of the CA is outside the scope of this study, continuous research on this is required. Reeves (2017) stated that these devices present challenging tasks in the computer-supported cooperative work and HCI areas because they throw fundamental questions regarding "cooperative interaction design

between heterogeneous human-agent groups" and "interaction between humans and machines."

The purpose of this study was to investigate how the combination of conversational agent and TV changes the user experience, using objective and subjective methods. Physiological measurements are useful for collecting objective data through the analysis of physiological recordings during interactions and identifying changes in user emotions. Although watching TV tasks are similar, the emotional responses of users during the CA interaction and the RCU interaction were opposites of each other. The results of physiological measurements are meaningful in that they can provide a more detailed and richer interpretation of the emotions and moods experienced during human-computer interactions. These results can also be used as a reference for TV viewing scenarios or interaction design for CA-equipped products. It is also interesting that participants evaluate the interaction with CA as more attractive than that with RCU even though they feel CA is inferior to RCU; the emotional quality contributed much more to the overall experience and satisfaction. Through this study, I confirmed that CA appeals more to user emotions and has a more positive impact. However, it is essential that a practical perspective is maintained in that CA is mainly adopted in everyday products used in daily life and usability is one of the most important attributes of such products. The physiological measurement results are considered to be useful enough to derive specific data on the user experience process, moving away from subjective assessment methods that typically indicate only a single direction, positive or negative. I hope that this study will contribute to providing meaningful findings in research on interactive technology and user experience mediated by conversational agents.

Chapter 6
Experiment 3. Long-term TV UX research with day reconstruction method (DRM)

6.1 Introduction

In recent years, users have demanded both “work-related and fun activities” (Thong et al., 2006). The ability to meet these demands requires factors other than satisfying users’ expectations of achieving their goals (Oliver et al., 1997); it requires that providers ensure a high level of user satisfaction by enabling users to experience pleasure and entertainment. Shin (2009), who studied the user experience of interactive TV using the technology acceptance model, identified perceived usefulness and perceived enjoyment as major factors in users’ acceptance of technology. Non-utilitarian products such as TV may require more enjoyment and relaxation qualities than pragmatic qualities. Weniger (2010) defined TV as a hedonic IT system in an Internet Protocol TV (IPTV) study and identified enjoyment as an important factor in users’ perception and adoption process related to TV.

Another studies on TV user experience (Bernhaupt and Pirker, 2013; Drouet and Bernhoupt, 2016) also determined that aesthetics and emotion were major factors related to user experience for interactive TV. Venkatesh (2000) defined enjoyment

as an antecedent of ease of use, and Norman (2004) also argued that emotional satisfaction plays a role in enhancing usability. Ease of use and enjoyment are connected because if a user has fun, then he/she will continue to use the technology, and as a result, use it more easily and enjoyably. As TV evolves into interactive media, a number of technologies surrounding TV is also changing. However, despite these changes, most TV interactions are handled by remote control, which does not accommodate the new and interactive features. Recently, CA, one of the technologies combined with TV, has been discussed as one of the alternatives that can accommodate functions that are difficult to control with RCU. The combination of CA with TV means that CA is used daily, not just occasionally. In this sense, the CA experience is different from one-time or occasionally used products.

Continuous and repetitive use experiences may provide different user experiences and affect short-term evaluations. In the 5 weeks Apple iPhone study, Karapanos et al. (2009) observed changes in the user experience over time. Kujala et al. (2012) conducted a longitudinal study to confirm the relationship between expectations and experience. Expectation affects the early stage of use, but as the duration of use increases, experiences that accumulate through usage have a greater influence on intention to use than expectations.

As shown in the study by Chitturi and colleagues (2008), the hedonic quality increases over time and delight affects customer loyalty. Walsh et al. (2014) argued that a positive long-term user experience positively contributes to customer loyalty. Blaynee et al. (2016) conducted a 2-year user observational study on the development of an incident report system and found that the initial user experience attributes were weakened and other attributes were strengthened over time. Through their research, they discovered the core needs and obstacles faced by people who are

at least 65 years old and are struggling with digital interface usage. According to the results, the authors proposed acceptable VR technology for elderly people. The advantage of longitudinal studies is that user-behaviour research can predict the acceptance of technology and the continuous intention to use by analyzing temporal patterns such as changes in experience over time. In this study, I explore how the CA interaction affects the user experience and how this experience differs between individuals by comparing CA and a remote control unit (RCU), which is a conventional TV controller in daily usages. The purpose of this study is to explore the user experience when users watch TV while interacting with CA and RCU, and whether long-term user experiences are related to use satisfaction and continuous intention to use.

6.2 Method

In this study, the day reconstruction method (DRM, Kahneman et al, 2004) was conducted to examine the TV experience in daily life. DRM is a method in which an individual records important anecdotes over time, recalling the day or the day before and evaluating the place where events occurred, the person with whom events occurred, the contents of the activity, and the emotional state experienced at that time. The advantage of DRM is that people can systematically reconstruct their daily experiences through anecdotal records over time. The measurement period of user experience was adapted from the previous study by Bhattacherjee and Premkumar (2004). According to Bhattacherjee and Premkumar, users' evaluations of IT service usually have a turbulent period and then stabilize after a month.

Based on these previous studies, the duration of the experiment was one month (4 weeks). Prior to the study, a pilot test of the measures and procedures was conducted.

The wording of the items was reviewed and modified based on the results of the pilot test obtained from 4 users. From November to December 2018, 20 households were asked to record their experiences with watching and controlling their TV for 4 weeks. Thirteen households used a TV equipped with CA. They had no prior experience of using the CA interaction with their TV. Seven households were selected to manipulate the TV with a remote control, which was their typical practice. The most common number of participants per family was four (45%), and the average number of members per participating family was 3.6. The average age of the participants was 39.65 years.

For 4 weeks, participants watched TV for at least one hour each day while manipulating the CA. After watching TV, they clicked a link that they received via messenger and participated in an online survey asking about their TV experience that day. The survey was constructed with an online form provided by NAVER, which is the largest portal in Korea. The questionnaires asked about the time spent watching the TV, the audience watching TV, the additional devices used when watching TV and users' experience and satisfaction. The survey response time was set from the moment when the link was sent to the next morning at 12 PM. Four hundred ten responses were collected from 20 participants for 28 days. Participants responded to additional questions about intention to use and overall satisfaction for a week, in addition to the daily reports. The experimental results were analyzed by user-based and data-based methods. The user-based analysis was performed according to the number of participants (N = 20) to collect the user profiles of participants. The data-based analysis was performed on the 410 responses collected over 28 days. A statistical analysis was performed to examine changes in user moods and user experience over time and to compare the results from the two interfaces: CA and RCU.

Before commencing the evaluation, participants answered 8 items asking about their expected satisfaction. Afterwards, participants rated their moods based on 10 items and reported their overall satisfaction after watching TV. The 10 questionnaire items were modified according to the study by Kahneman and colleagues (2004); 5 items assessed positive affect, such as happy and interested, and 5 items assessed negative affect, such as bored and frustrated. Table 6.2.1 shows the results of the reliability and factor analysis of the questionnaire. The factor analysis revealed two main factors, negative and positive, with reliability values of .855 and .850, respectively. Factors 1 and 2 accounted for 42.29% and 24.48% of the variance, respectively.

Table 6.2.1 The results of the factor analysis of affect.

	Factor I	Factor II	Cronbach Alpha
q8 Frustrated/annoyed	.896	-.205	
q4 Anxious	.885	-.029	
q6 Depressed / Blue	.865	-.073	.855
q2 Board	.754	-.114	
q10 Tired	.494	-.067	
q3 Happy	-.183	.857	
q9 Warm/Friendly	-.051	.850	
q1 Competent/Capable	.267	.807	.850
q5 Interesting	-.285	.721	
q7 Comfortable	-.426	.683	
Eigen value	42.289	24.482	
Kaiser-Mater-Olkin Measure of Sampling Adequacy			.814
Bartlett's Test of Sphericity		Approx. Chi-square	629.486
		df	45
		Sig.	.000***

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

In addition to reporting their own TV experience each day, participants responded to four questions once every seven days asking about their intention to use the technology. Intention to use was assessed with a questionnaire comprising four items from the study by Kujala et al. (2012). Confidence in these questions was assessed through a reliability analysis (Cronbach's alpha = .849).

6.3 Results

Participants were asked to rate their moods by answering eight questions after watching TV each day for 4 weeks. For the positive mood questions, participants answered that they felt comfortable (M = 4.69), interested (M = 4.48) and happy (M = 4.36). For the negative mood questions, participants reported that they were tired (M = 3.41), anxious (M = 2.69) and bored (M = 2.65).

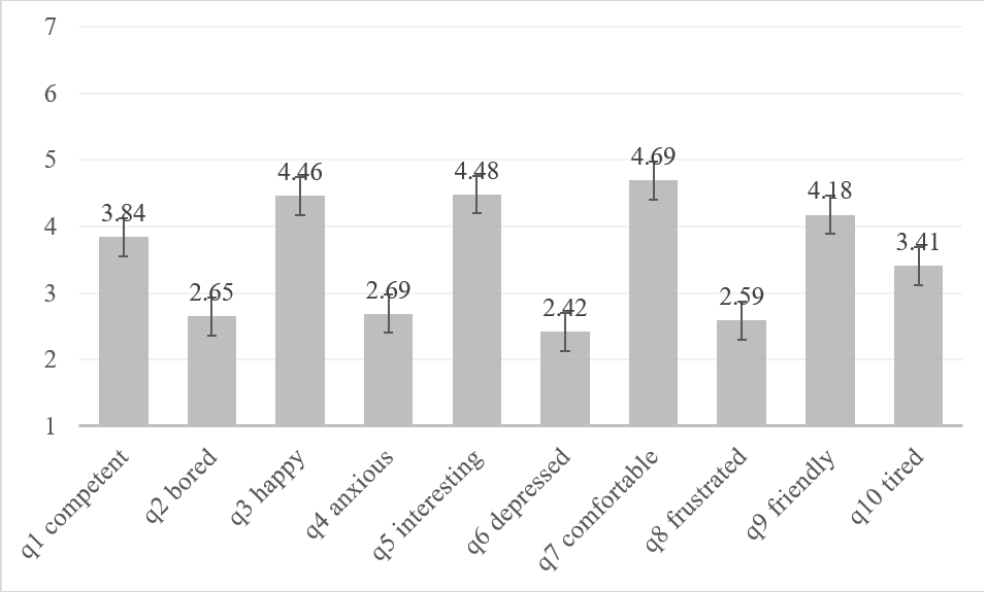


Figure 6.3.1 TV viewers' self-reported moods over 28 days.

Table 6.3.1 shows the results of a repeated-measures ANOVA designed to analyse whether participants' moods after watching TV varied according to the interfaces. The results showed a significant difference according to the number of days ($F = 2.281$, $p < .01$) and a correlation between the interface and days ($F = 2.564$, $p < .01$). Interactions between days, interfaces and moods did not appear.

Table 6.3.1 F-values from repeated measures ANOVA of user's daily moods after interacting with each interface (RCU and CA) for 28 days

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Days	Sphericity Assumed	138.567	27	5.132	2.281	.005**
Days * Interface	Sphericity Assumed	155.756	27	5.769	2.564	.002**
Days * Interface * Moods	Sphericity Assumed	389.339	243	1.602	.869	.892

*Level of significance. *** p value < .001, ** p value < .01, and * p value < .05

The differences in the participants' moods between interfaces over time is shown in Figure 6.3.2. This graph shows users' moods for 28 days. The daily emotional state of the RCU users did not change substantially, but the CA users' emotional states showed relatively large variation. The reason for this change is that because of the small number of participants, one or two opinions may affect the change of results. It is predicted that the degree of emotions (false, fun, etc.) caused by CA is greater than that caused by RCU. The participant's diary analysis confirms this implication. Participants are annoyed by not being able to understand what they are saying, or they are comfortable with being able to operate the TV while doing other things. Thus, participants using CA were influenced to a greater extent by the interaction.

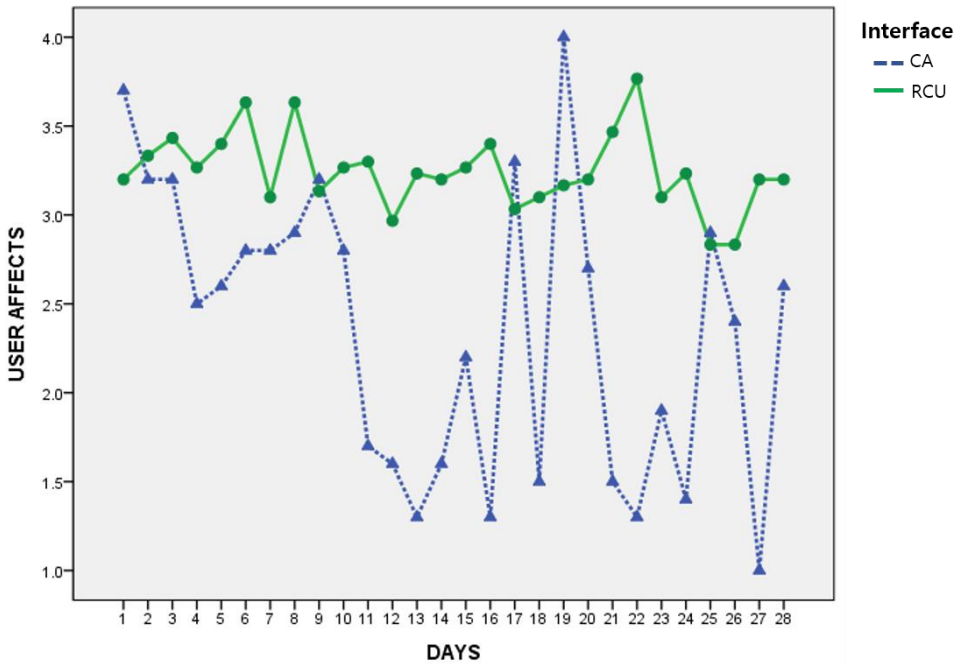


Figure 6.3.2 Comparison of self-reported affects between users of the CA interaction and the RCU interaction for 28 days.

A significant difference between the expected satisfaction measured before the experiment and the satisfaction experienced over 4 weeks was observed. In Figure 6.3.3, the left graph shows expected satisfaction, and the next four graphs show experienced satisfaction for 4 weeks. The results reflect repeated use satisfaction. Overall, the experienced satisfaction was higher than the expected satisfaction, and the experienced satisfaction increased continuously over time.

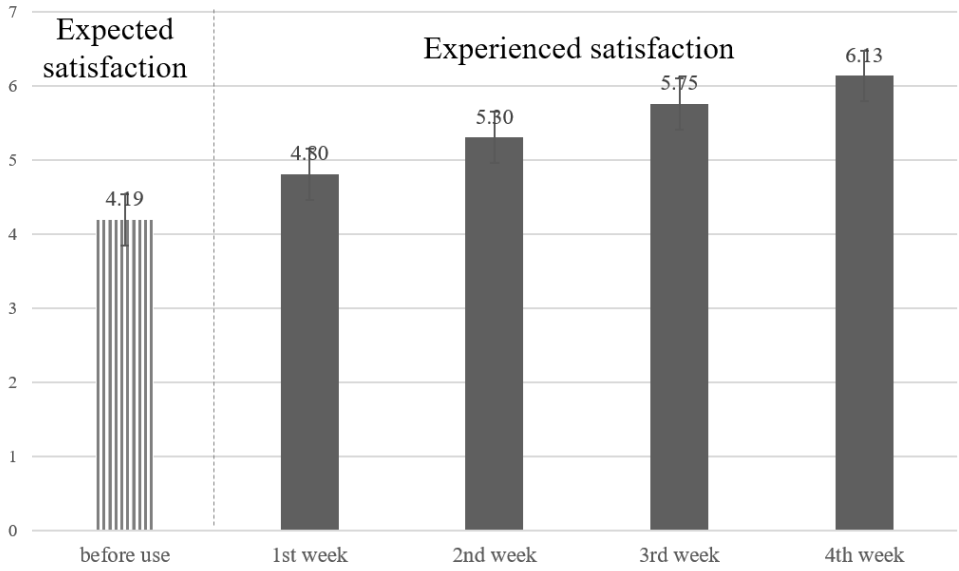


Figure 6.3.3 The results of TV viewers' expected satisfaction and experienced satisfaction.

Based on the expected satisfaction (M = 4.187), the differences between the expected and experienced satisfaction are shown in the upper part of Table 6.3.2. All weeks showed significant differences as follows: 1st week t (19) = 5.328, p < .001; 2nd week t (19) = 4.201, p < .001l; 3rd week t (14) = 5.497, p < .01; and 4th week t (15) = 9.061, p < .001. The largest difference was observed in the 4th week (mean difference = 1.947). However, the difference between the interfaces (CA and RCU) was not statistically significant. The bottom of Table 6.3.2 shows the comparison of results for intention to use and expected satisfaction. The difference in intention to use and expected satisfaction was also statistically significant in all weeks (1st week t (16) = 4.428, p < .001; 2nd week t (15) = 2.509, p < .05; 3rd week t (11) = 3.344, p < .01; 4th week t (13) = 3.905, p < .01). The most significant difference was detected in the last week (mean difference = 1.188).

Table 6.3.2 Results from the one-sample t-test of satisfaction for 4 weeks

		Test value (Expected satisfaction) = 4.187					
		t	df	Sig. (2-tailed)	Mean difference	95% confidence interval of the difference	
						Lower	Upper
Experienced satisfaction	1st week	5.328***	19	.000	.61693	.3746	.8593
	2nd week	4.201***	19	.000	1.11443	.5592	1.6697
	3rd week	5.497***	14	.000	1.5649	.9544	2.1755
	4th week	9.061***	15	.000	1.94693	1.4889	2.4049
Intention to use	1st week	4.428***	16	.000	1.01888	.531	1.5067
	2nd week	2.509*	15	.024	.73487	.1106	1.3592
	3rd week	3.344**	11	.007	.91717	.3134	1.5209
	4th week	3.905**	13	.002	1.188	.5308	1.8452

*Level of significance. *** p value < .001, ** p value < .01, and * p value < .05

Repeated-measures ANOVA was conducted to confirm the changes in intention to use over time (Figure 6.3.4). The left graph in Figure 6.3.4 shows intention to use recorded for four weeks. The intention to use showed slightly different results from the results for experienced satisfaction. User satisfaction increased steadily from the 1st to 4th weeks. However, the intention to use decreased in the 2nd week and then increased again in the 3rd week. As shown in the graph, a slight decrease was observed between the 1st week (M = 5.21) and the 2nd week (M = 4.92) and a slight increase was observed between the 3rd week (M = 5.10) and the 4th week (M = 5.37).

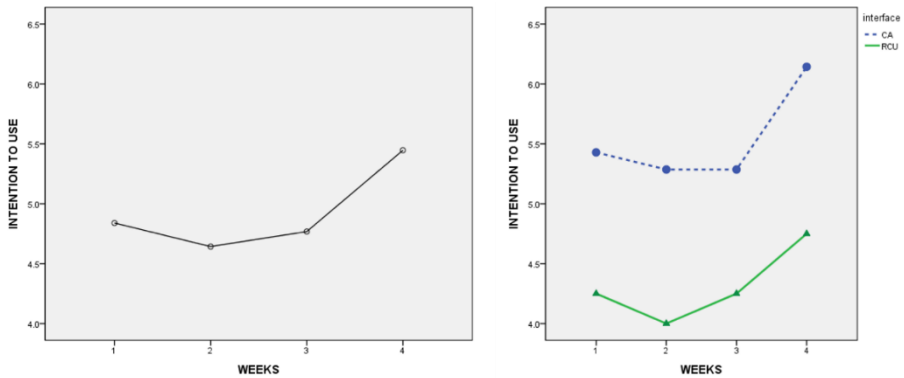


Figure 6.3.4 Changes in intention to use over 4 weeks (left panel). Changes in intention to use CA and RCU over 4 weeks (right panel).

The graph on the right side of Figure 6.3.4 shows the results for the intention to use the two interfaces. The overall trend of changes was slightly decreasing and increasing for both CA and RCU. The increase in the intention to use for over three weeks seems to be due to increased user satisfaction. To find out why intention to use increased again in 4th week, text analysis was conducted on the users' diaries.

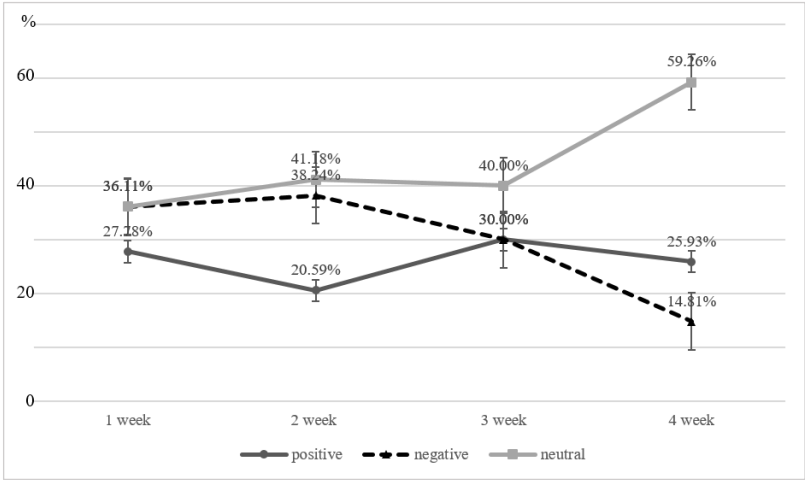


Figure 6.3.5 The user's comments about TV related experience for 4 weeks

The table above shows 4 weeks user comments recorded by TV viewers who watched TV in the home using CA or RCU interfaces. Many comments were related to the CA interaction rather than the RCU interaction. But the majority of comments did not directly indicate the interfaces in watching TV. The user comments appeared in the first week and the second week had a lot to say about the voice recognition errors and operational errors of CA. In the second week, positive comments about CA decreased slightly and negative comments increased compared to the first week. The user comments in the first and the second weeks were more negative than positive. The result of the third week showed the positive comments and the negative contents were almost same. At the forth week, the rate concerning CA in emotional aspects decreased overall.

In the other hand, most comments were neutral. Neutral comments refer to behaviors-directed reports such as "I asked the weather information to the CA," without showing any positive or negative emotions associated with watching TV. At the same time, user satisfaction and the continuous usage intention increased at the forth week. It implies that the users have learned the various usage errors that occurred while using CA first, and then use it within the controllable range. In other words, they are somewhat disappointed with the quality of CA compared to expectation, and use it to some extent to recognize the quality and adjust their usage range according to the quality level of CA. This implication matches to some extent in the user comments. Many user comments on the first and the second weeks were confused or annoyed. Contrary, user comments in the third and the fourth weeks showed that users use CA in necessary and take advantage of CA.

Table 6.3.3 The users' daily comments about TV related experience

Positive user comments	Negative user comments
"It was so easy to control the TV with voice while feeding my baby milk. Voice recognition is essential when caring for a baby."	"I was surprised by the sudden change of TV channels while talking with my family. Maybe Genie misunderstood."

“It is really necessary to find the remote control function by voice when watching TV alone.”	“I tried to find the TV content by voice, but I gave up because of poor recognition accuracy.”
“It's comfortable to turn the TV on and off while lying down.”	“While watching TV today, I tried to turn off the TV because I was embarrassed that Gennie suddenly played music.”
“My children seem to think it is easy to search TV contents by voice command than the remote control.”	“I asked Gennie to turn the channel to 11, but Genie did not recognize it. I asked again “MBC,” and then Gennie recognized my voice.”
“I was able to set the time schedule to turn off the TV after 30 minutes, by asking Gennie. Very comfortable!”	“Voice operation is still difficult to control VOD content.”
I asked Gennie, "How's the fine dust today?" Gennie responded. "The fine dust level is normal."	“I asked Gennie to show the 5th TV program on the content list, but it didn't get it well. I wish Gennie have the ability to understand context more.”

In the previous studies, it was found that the more positive the experience after use compared to the expectation before use, the greater the impact on user satisfaction (Deng et al., 2006; Solomon, 1996; Oliver, 1980; Oliver & Wayne, 1988; Yi, 1990; Bhattacherjee & Premkumar, 2004). According to Kujala (2012), etc., satisfaction also increases over time. Therefore, it can be predicted that the change in the graph of continual usage intention results from increasing satisfaction.

The intention to use was slightly different between the two interfaces. The intention to use CA was slightly decreased until the 3rd week and increased significantly at the 4th week. The intention to use RCU showed the lowest value in the 2nd week and then increased again in the 3rd week. Overall, the intention to use CA (M = 5.06) was slightly higher than the intention to use RCU (M = 4.96), but the difference was

not statistically significant. These results may be attributed to the number of participants (RCU = 7 and CA = 13), which may have been too small to obtain significant differences. I merged the intention to use data collected in the 1st, 2nd, and 4th weeks, which showed similar trends, and repeated the analysis to address this question. The combined results for intention to use showed slightly significant differences between CA and RCU (t (37) = 2.254, p < .05) (Table 6.3.4). Based on this results, a larger number of subjects should be investigated in related studies. This result will be used as a reference and will not be used as formal data in the present study.

Table 6.3.4 The results of the one-way ANOVA explain the difference in intention to use between the two interfaces (CA and RCU) in the 1st, 2nd, and 4th weeks

	F	Sig.	t (df)	Sig. (2-tailed)	Mean difference	Std. error of the difference	95% confidence interval of the difference	
							Lower	Upper
Intention to use at 1, 2, and 4 weeks	5.890	.02	2.254 (37)	.030*	1.075	.476	.108	2.041

*Level of significance. *** p value < .001, ** p value < .01, and * p value < .05

The relation between experienced satisfaction and intention to use did not indicate a statistically significant difference between the two interfaces. Therefore, further studies analyzing the effect of satisfaction on intention to use were conducted without distinguishing the interfaces. A linear regression analysis was conducted to determine whether experienced satisfaction affected intention to use in the same manner as in the first experiment (Table 6.3.5). Experienced satisfaction, but not expected satisfaction, exerted a significant effect on intention to use (t (19) = 3.335,

p < .01).

Table 6.3.5 The results of the linear regression analysis explaining the effect of experienced satisfaction on intention to use

		B	Standard Error	β	t	sig.
Intention to Use	Satisfaction	.846	.254	.629	3.335	.004**
	F = 11.123 (p < .01), R^2 = .396, D-W = 1.873					

*Level of significance. *** p value < .001, ** p value < .01, and * p value < .05

Table 6.3.5 shows the effect of experienced satisfaction on intention to use in each week. The results obtained during the 2nd week showed a statistically significant effect on intention to use (F (3, 34) = 9.662, p < .01; t (19) = 3.108, p < .01). The results from the 1st and 4th weeks revealed nonsignificant, but nearly significant, effects of experienced satisfaction on intention to use. Expected satisfaction also did not exert a significant effect on intention to use.

Table 6.3.6 The results of the linear regression analysis explaining the effect of experienced satisfaction on intention to use for 4 weeks

	F	R^2	Satisfaction	B	S.E.	β	t	sig.
Intention to Use	3.609	.194	1st week	.804	.423	.440	1.900	.077
	9.662**	.408	2nd week	.723	.232	.639	3.108	.008**
	.962	.088	3rd week	.238	.243	.296	.981	.350
	4.740*	.283	4th week	.681	.313	.532	2.177	.050

*Level of significance. *** p value < .001, ** p value < .01, and * p value < .05

6.4 Conclusion

In this study, it was explored how the experience of using CA while watching TV affects users' satisfaction and intention to use this technology in daily life. Twenty users participated in this experiment and evaluated their experiences and satisfaction daily for 4 weeks. Intention to use was also measured weekly. The most frequent positive feeling while watching TV was comfort. Tired and bored were the most frequently reported negative feelings. CA users and RCU users reported different moods and experiences over time while watching TV. The flow of the mood graph showed that the daily emotional state of the RCU users did not change substantially, but the CA users' emotional states showed relatively large variations. According to the results of accumulated experience, user satisfaction increased over time compared to expected satisfaction before use. The mean value for user satisfaction was higher for CA than for RCU, although the difference was not statistically significant. Intention to use was measured weekly.

The results of daily evaluated user satisfaction and weekly evaluated intention showed slight differences over time. Intention to use decreased in the 2nd week and then increased thereafter. In particular, the intention to use CA decreased slightly until the 3rd week but increased again during the 4th week. Statistically significant results were not obtained for the difference in intention to use between the interfaces. Experienced user satisfaction exerted a direct effect on intention to use. The results were similar to the results from the study 2. Expectation did not show any effect on intention to use. A summary of the findings from experiment 1 and experiment 2 is presented below. RCU was rated higher than CA in pragmatic quality dimensions. CA was rated higher than RCU in hedonic quality dimensions. In both experiments, experienced satisfaction exerted a direct impact on intention to use. The effect of

expected satisfaction before use was not insignificant or non-existent. The user's perceived experience exerted a significant impact on the intention to use both the old and new interfaces.

According to the results of 4 weeks self-diaries, a significant difference in the correlation between time and the interface was observed. Experienced satisfaction continuously increased over time compared to expectation before use. In particular, the results obtained from the 1st week and the 4th week were different and outside the range of error. Experience accumulated and affected user satisfaction over time. A correlation between satisfaction and intention to use was also observed. However, the trends for satisfaction and intention to use did not show the same results. Satisfaction showed a gradual increasing trend from the beginning of the experiment, but intention to use initially decreased and subsequently increased. Based on these results, satisfaction is not immediately linked to intention to use and a time disparity may exist between the two factors.

There were some difficulties in controlling such as interfaces in the field study. The original intention was to distinguish CA users and RCU users. However, CA users actually used RCU and CA together. This situation may have the advantage that the results better reflect actual experiences. At the same time, the results are limited by the lack of a clear difference between the two experiences. The user diaries indicated that CA was still used as a complementary method, not as a main controller, when watching TV. When users watched TV, they still used RCU for switching channels or menu navigation. CA was used as a control for simple commands such as "Turn on the TV" and alternatives to manipulating complex features (e.g., text input). Thus, CA has not yet evolved sufficiently to meet all user needs and instead serves as an assistant. Most CA users experienced a relatively frustrating process due to frequent

communication errors while attempting conversations at the beginning of the experiment. A graph showing a decrease in intention to use for a period highlights this point, implying that CA experience may initially be disappointing due to low performance relative to expectations.

According to a related study (Forlizzi and Battarbee, 2004), the experience of users who do not have previous experience with a new technology differs from experienced users. The authors asserted that users who have little experience with a new technology are very disappointed after use. Experienced users, however, are aware of the limitations of the technology, and therefore they are less disappointed. Interestingly, intention to use increased in the fourth week. Some explanations for this increase are found in the participants' diaries. Users indicated that the voice interaction was helpful when they were unable to use their hands or were focusing on watching TV while doing housework. Additionally, participants had different thoughts about RCU and CA. They evaluated CA to be 'worthy' or 'interesting' if it worked properly, whereas they viewed RCU as an inanimate thing. After experiencing disappointing moments, users may establish a mental model for CA and start to mainly interact with CA with simple and error-free commands such as "Go to channel number 5".

Despite its contributions, the present study has limitations. In the laboratory experiment, users were able to operate the TV with only one interface per session. For this reason, some users may not have been able to complete the task in the CA session. This result probably explains why the mean value of PQ for CA was lower than for RCU. In the long-term experiment, I had minimal control when acquiring real data from users. Many cases of missing data were identified. In the weekly evaluations, which were collected once a week for 4 weeks, only 13 valid data points

were available. In addition, the evaluation was not observed directly, and uncontrolled variables from the situation and the environment may have confounded the results. The actual situation that the CA users mentioned above, namely, the concurrent use of CA and RCU, may also have affected the results. RCU and CA usage were not completely controlled, and CA users used both CA and RCU simultaneously. If the results for the daily measurements of mood had shown no difference between the two interfaces, then the CA may not have had any effect on the user experience.

However, the daily mood recorded by the users showed a difference between the interfaces, and I concluded they had different experiences with the interfaces. Another limitation is the duration of the experiment. The duration of 4 weeks was determined based on previous research (Bhattacherjee and Premkumar, 2004), but it is still unsatisfactory because the results obtained at the 4th week do not confirm whether user satisfaction remains constant after the rapid increase from the 3rd week. The period must be extended to 6 to 8 weeks to record observations after 4 weeks and conclude that experience stabilizes within 4 weeks (Kujala et al., 2012). In future studies, the number of households must be increased to evaluate daily data.

A direct comparison of the results from the lab test with the daily field data may not be appropriate. However, two evaluations were conducted based on the merits, rather than the disadvantages, of each method. Laboratory testing is important to clearly compare the differences between the two interfaces and allows researchers to determine which UX factor is more likely to affect overall evaluation about each interface. The advantage of field testing is that users experience real-world satisfaction and intention to use while using a technology naturally in a real-life environment. Additionally, researchers can determine whether satisfaction and

acceptance change over time.

In this study, it was investigated that the user experience of a CA-equipped TV to assess the intention to use CA, which is evolving and rapidly penetrating our daily lives. The intention to use CA was based on the user's experienced satisfaction, which was confirmed in both short-term and long-term studies. Despite the increasing prevalence of CA, few studies have investigated CA from the perspective of the user and developed appropriate mental models. The construction of an accurate mental model of how users interact with CA is urgently needed and important. Accurate feedback on the level of user satisfaction and continued use of the system may be related to persistent intention to use mediated by satisfaction.

The following points may be considered to improve user satisfaction with CA based on the results from this study. First, CA manufacturers must be able to quickly and accurately recognize the level of CA interaction and help users have optimized experiences. As soon as users have an accurate grasp of the technology, their disappointment decreases and they are more likely to use the technology as their experience accumulates. At the same time, the emotional quality of CA must be strengthened. According to Partala et al. (2004), the emotional content of the CA voice can affect the user's emotional experience.

Another study showed that adapting non-verbal behaviours, such as tone and pitch, which are observed in human interactions, may be used to enhance human-CA interactions (Foster, 2007). The ability of CA to emotionally appeal to users is one of the conditions that must be considered in product development in the future. Since emotional quality represents a different dimension of usability, non-utilitarian factors, such as fun and pleasure, must be enhanced, rather than only increasing functionality or efficiency. The fact that interacting with voice is stimulating and emotionally

enjoyable must be considered an integral part of designing CA applications. Currently, voice recognition technology has been added without changing the existing TV viewing context. The interface design must be altered after considering functional aspects and emotional aspects of the human-CA interaction. This study confirms that experience affected user satisfaction in both the short-term study and the longitudinal study and that satisfaction based on experience directly affected intention to use. In addition, user satisfaction with new technology may be low at the beginning of use, but accumulated experience in daily life can positively affect satisfaction and intention to use. Therefore, this study is important because it discovered the necessity of a long-term method for evaluating the user experience.

Chapter 7
The validation of UX model framework

7.1 The conceptual model of interactive TV UX

In this chapter, Structural Equation Modeling (SEM) examines the suitability of measurement variables for the interactive TV UX framework and the causal relationship among the main factors for TV UX. The reasons for applying the structural equation model are as follows: The studies performed in this paper are about the user experience and its interaction with TV. Very few studies on interactive TV evaluate and measure the entire range of processes before and after the user experience.

The first reason for doing the above is to confirm the validity of applying the measurement scale presented in the previous study on the interactive product to the interactive TV UX framework. The second reason is to confirm the causal relationship between the expected and experienced UX quality factors, and the overall satisfaction and continuous intention to use. This is to obtain an empirical evidence that actual experience is necessary for acceptance of new technologies. Confirmatory factor analysis and path analysis were performed by using the structural equation model. Confirmatory factor analysis verifies the adequacy of the variables used in the measurement model, and confirms that there is no problem when the previously validated measurement tool is applied to the interactive TV UX study. The latent variables defined in this paper are expected satisfaction, perceived PQ, perceived HQ, perceived satisfaction, effects, and intention to use. In addition, each latent variable has sub-measured variables. Through this, sub-variables that are

not suitable for the interactive TV UX research model are removed and variables suitable for the interactive TV UX research model are summarized. After verifying the research model, structural equation modeling was conducted to confirm the causal relationship between the UX quality factors through path analysis.

A total of 82 data points were applied to the structural equation model. The usually recommended subjects for structural equation models are more than 200 people (Boomsma, 1982). However, it is not suitable in this case because this is an empirical study, which makes it difficult to gather subjects for interactive TV UX. Nevertheless, some results through interactive TV UX research were found to be useful as the model suitability was high. First, the research hypothesis on the interactive TV user experience is defined as shown in Figure 7.1.1 below.

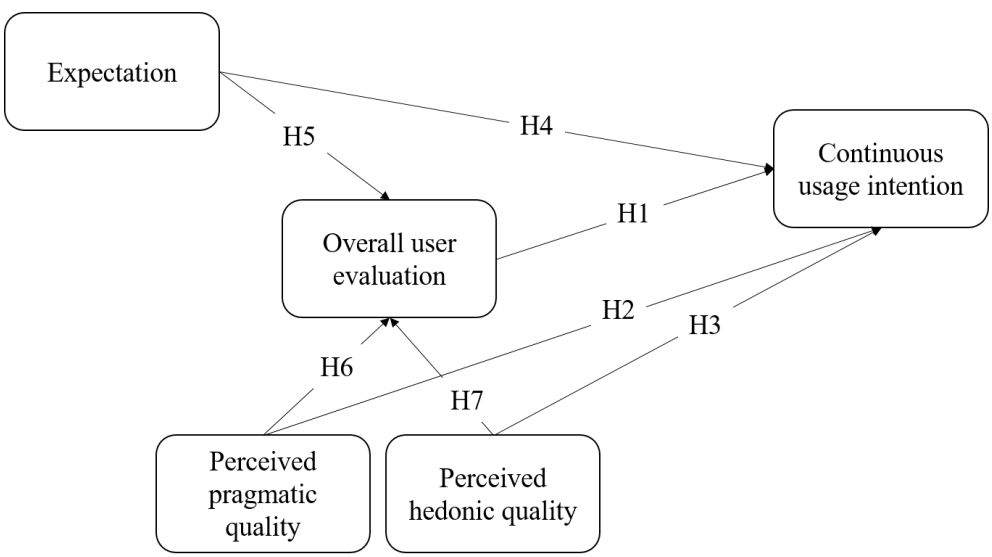


Figure 7.1.1 Hypothesis model about interactive UX framework (original model)

Figure 7.1.1 is the hypothesis model of interactive TV UX based on the related studies discussed in the chapter 2. The characteristic of this model is that the

antecedent of continual usage intention technology is based on the perceived user satisfaction (the result of overall user evaluation) (H1). Expected satisfaction will also affect intention to use (H4). The perceived user satisfaction is influenced by the actual user experience and the expected satisfaction (H4). In addition, the users are experienced in two dimensions – pragmatic quality (H6), and hedonic quality (H7). These perceived qualities will also affect continual usage intention (H2, H3).

The descriptions for the latent and measured variables used in the model are show below (see Table 7.1.1). Among the measured variables, “frustrating” (BEF1) and “spend a lot of time” (BEF3) of the expectation category were used as reversed values to ensure consistency in the results.

Table 7.1.1 Descriptive statistics

Category	Attributes	Mean	S.D.	Skewness	Kurtosis
expectation	BEF1_frustrating	4.54	1.484	-.317	-.476
	BEF2_easy to use	5.04	1.319	-.698	.060
	BEF3_spend a lot of time	3.99	1.536	.063	-.705
	BEF4_meet requirement	4.85	1.208	-.530	-.260
	BEF5_enjoy	4.48	1.484	-.335	-.860
	BEF6_fun	4.46	1.450	-.560	-.233
	BEF7_rewarding	3.44	1.415	-.235	-.861
	BEF8_satisfaction	4.60	1.422	-.414	-.900
Affects	Valance	5.21	1.844	-.168	-.865
	Arousal	4.61	1.797	-.098	-.131
Pragmatic Quality	PQ1_technical - human	3.22	1.663	.500	-.981
	PQ2_complicated - simple	4.20	1.644	.055	-.993
	PQ3_impractical - practical	4.11	1.764	-.088	-1.154
	PQ4_cumbersome - straight forward	3.76	1.775	.203	-1.082
	PQ5_unpredictable - predictable	4.51	1.687	-.219	-1.016

	PQ6_confusing - clearly structured	4.29	1.486	.149	-.729
	PQ7_unruly - managable	4.70	1.668	-.400	-.827
Hedonic Quality-Identification	HQI1_isolating - connective	4.23	1.443	-.216	-.322
	HQI2_unprofessional - professional	3.68	1.323	.216	-.120
	HQI3_tacky - stylish	4.11	1.286	-.210	.017
	HQI4_cheap - premium	4.06	1.318	-.181	.066
	HQI5_alienating - integrating	4.12	1.180	.035	.216
	HQI6_separates me - bring me closer	4.33	1.287	-.359	-.135
	HQI7_unpresentable - presentable	4.28	1.550	-.240	-.801
Hedonic Quality-Stimulation	HQS1_conventional - inventive	3.74	1.910	.104	-1.279
	HQS2_ unimaginative - innovative	3.63	1.629	.193	-.887
	HQS3_cautious - bold	4.00	1.305	.171	-.119
	HQS4_convservative - innovative	3.84	1.575	-.042	-.582
	HQS5_dull - captivating	3.80	1.503	-.128	-.586
	HQS6_undemanding - challenging	4.06	1.550	-.328	-.992
	HQS7_ordinary - novel	3.29	1.495	-.019	-.866
APPEAL	ATT1_unpleasant - pleasant	3.72	1.493	-.005	-.664
	ATT2_ugly - attractive	4.29	1.060	.213	.839
	ATT3_disagreeable - likeable	4.35	1.148	-.283	.740
	ATT4_rejectable - inviting	4.57	1.361	-.293	-.208
	ATT5_bad - good	4.57	1.111	-.853	1.599
	ATT6_repelling - appealing	4.23	.998	-.409	2.102
	ATT7_discouraging - motivating	4.04	1.181	-.441	.174
Intention to use	INT1_continuous to use	4.23	1.460	-.366	-.758
	INT2_continuous to use from now	4.12	1.494	-.305	-.710
	INT3_recommend to friend	4.04	1.688	-.169	-.819
	INT4_use again in the future	4.48	1.657	-.565	-.729

7.2 Validation of interactive TV UX model framework

Figure 7.2.1 shows the relationship between the user's expectation before interaction, the overall user experience based on the perceived UX qualities produced by the interaction, and finally the continuous intention to use the system.

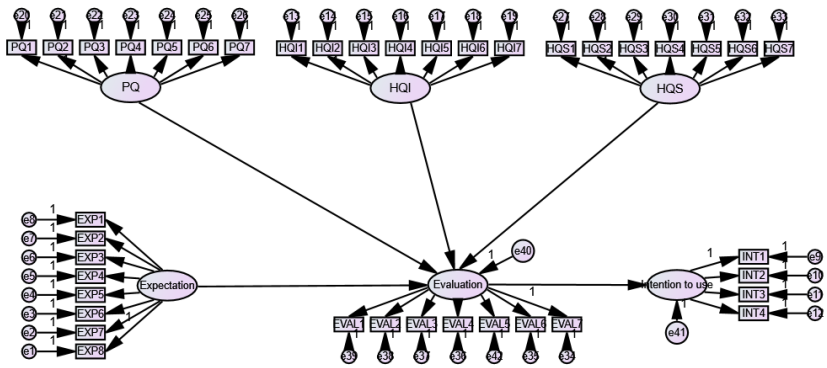


Figure 7.2.1 Original model of interactive TV UX attributes

This is a very large structural model that can be very complex to fit to a full structural equation model because each latent variable has lots of item. Actually the model-fit of this structure have showed very low and the error rates are also high. To increase model-fit, the structural equation model was established after dividing the experience process model and the quality process model affecting the user experience.

7.2.1 Confirmatory Factor Analysis of interactive TV UX process model

This is the result of factor analysis (Figure 7.2.2) of UX model which does not include the UX quality factors that form the user experience. This model consists

only the phase based UX factors – expectation, experience and intention to use – as latent variables.

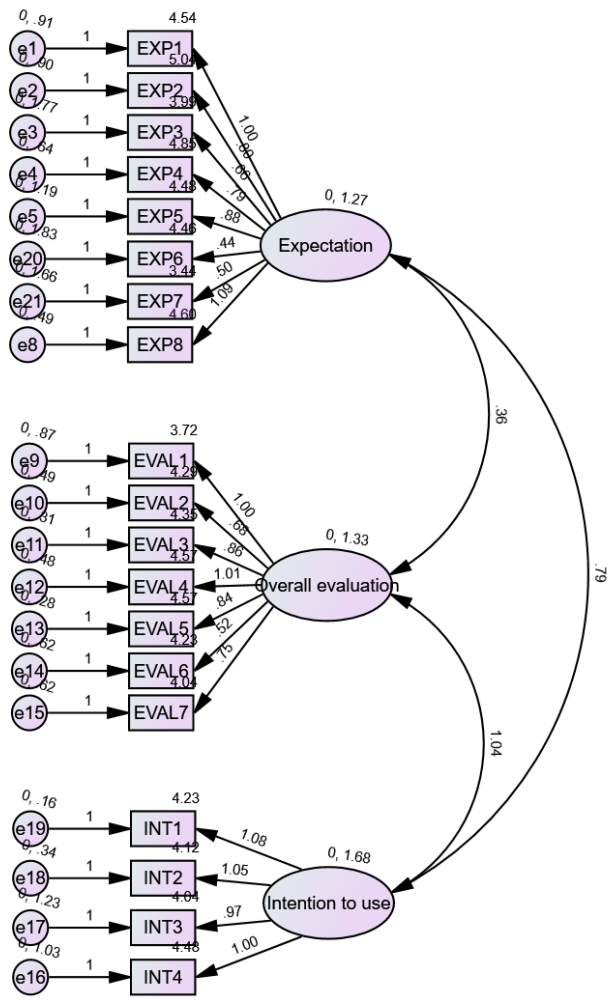


Figure 7.2.2 Original model of UX attributes

As a result of the original model measurement, the model fits close to the criterion of fitness, but it does not reach the goodness-of-fit yet ($\chi^2 = 285.707$ ($p < .001$), TLI

= .847, CFI = .867, RMSEA = .106).

Table 7.2.1 Regression analysis result for UX attributes (original model)

			Estimate		S.E.	C.R.
			B	*β*		
BEF1_frustrating	<---	Expectation	1	0.783		
BEF2_easy to use	<---	Expectation	0.799	0.704	0.124	6.47***
BEF3_spend a lot of time	<---	Expectation	0.684	0.517	0.149	4.582***
BEF4_meet requirement	<---	Expectation	0.778	0.747	0.112	6.937***
BEF5_enjoy	<---	Expectation	0.809	0.633	0.141	5.733***
BEF8_satisfaction	<---	Expectation	1.061	0.867	0.13	8.163***
ATT1_pleasant	<---	Overall evaluation	1	0.778		
ATT2_attractive	<---	Overall evaluation	0.679	0.745	0.095	7.172***
ATT3_likeable	<---	Overall evaluation	0.864	0.874	0.098	8.774***
ATT4_inviting	<---	Overall evaluation	1.008	0.86	0.117	8.597***
ATT5_ good	<---	Overall evaluation	0.84	0.878	0.095	8.829***
ATT6_ appealing	<---	Overall evaluation	0.522	0.608	0.092	5.651***
ATT7_ motivating	<---	Overall evaluation	0.751	0.739	0.106	7.106***
INT1_continuous to use	<---	Intention To Use	1.078	0.962	0.105	10.298***
INT2_continuous to use from now	<---	Intention To Use	1.055	0.92	0.108	9.763***
INT3_recommend to friend	<---	Intention To Use	0.97	0.748	0.131	7.406***
INT4_use again in the future	<---	Intention To Use	1	0.786		

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

Table 7.2.2 Fitness index of measurement model (original model)

Model	χ^2	df	TLI	CFI	RMSEA
	285.707***	116	0.847	0.867	0.106

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

In order to improve the model fit, “fun” (EXP6) and “rewarding” (EXP7) variables were removed from the expectation variables with estimates less than 0.5 in the Standardized Regression Weights. The results showed that there was a weak but positively significant causal relationship between the expectation before use and experience after use ($\beta = 0.34$, $p < .05$). The expectation before use also showed a significant correlation with the intention to use ($\beta = 0.79$, $p < .001$). There was a strong correlation between the overall evaluation and intention to use ($\beta = 1.04$ $p < .001$). The goodness-of-fit of the modified measurement model was found to be satisfactory as $\chi2 = 184.16$ ($p < .001$), TLI = .915, CFI = .928, and RMSEA = .085 (see Table 7.2.2).

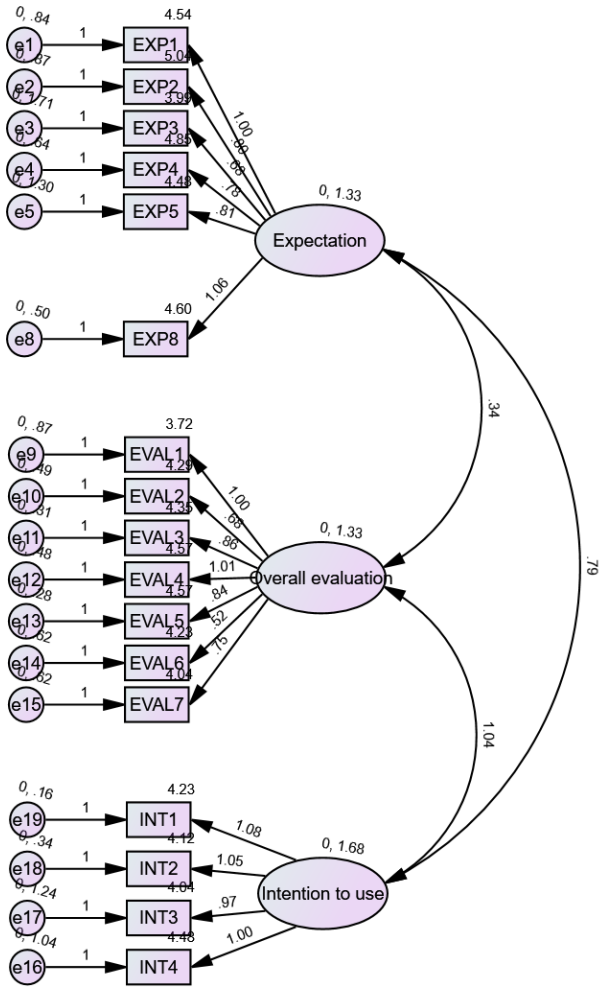


Figure 7.2.3 Modified model of UX attributes

Table 7.2.3 shows the results of the modification after removing the variables in order to meet the structural equation model criteria in terms of the expected satisfaction, overall evaluation, and continuous intention to use. Based on the factor analysis, it was found that all the sub-measurement variables, except fun and

rewarding, were significant.

Table 7.2.3 Regression analysis results (modified model)

Variables			estimate		S.E.	C.R.
			B	*β*		
Overall evaluation	<---	Expectation	0.238	0.253	0.115	2.07*
Intention To Use	<---	Overall evaluation	0.721	0.596	0.12	6.028***
Intention To Use	<---	Expectation	0.433	0.38	0.101	4.296***
BEF8_satisfaction	<---	Expectation	1	0.867		
BEF5_enjoy	<---	Expectation	0.763	0.633	0.125	6.124***
BEF4_meet requirement	<---	Expectation	0.733	0.747	0.096	7.653***
BEF3_spend a lot of time	<---	Expectation	0.644	0.517	0.135	4.775***
BEF2_easy to use	<---	Expectation	0.753	0.704	0.107	7.044***
BEF1_frustrating	<---	Expectation	0.942	0.783	0.115	8.163***
ATT1_pleasant	<---	Overall evaluation	1	0.778		
ATT2_ attractive	<---	Overall evaluation	0.679	0.745	0.095	7.172***
ATT3_ likeable	<---	Overall evaluation	0.864	0.874	0.098	8.774***
ATT4_ inviting	<---	Overall evaluation	1.008	0.86	0.117	8.597***
ATT5_ good	<---	Overall evaluation	0.84	0.878	0.095	8.829***
ATT6_ appealing	<---	Overall evaluation	0.522	0.608	0.092	5.651***
ATT7_ motivating	<---	Overall evaluation	0.751	0.739	0.106	7.106***
INT1_continuous to use	<---	Intention To Use	1	0.962		
INT2_continuous to use from now	<---	Intention To Use	0.979	0.92	0.062	15.847***
INT3_recommend to friend	<---	Intention To Use	0.9	0.748	0.097	9.291***
INT4_use again in the future	<---	Intention To Use	0.928	0.786	0.09	10.298***

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

The structural equation model confirmed that the expectation, experienced qualities and satisfaction, and intention to continuous use potential variables and sub-variables were selected appropriately for the interactive product UX study. This implies that it is suitable for use as a measurement scale for the study of other interactive system UX in the future.

Table 7.2.4 Fitness index of measurement model (modified model)

Model	χ^2	df	TLI	CFI	RMSEA
	184.16***	116	0.915	0.928	0.085

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

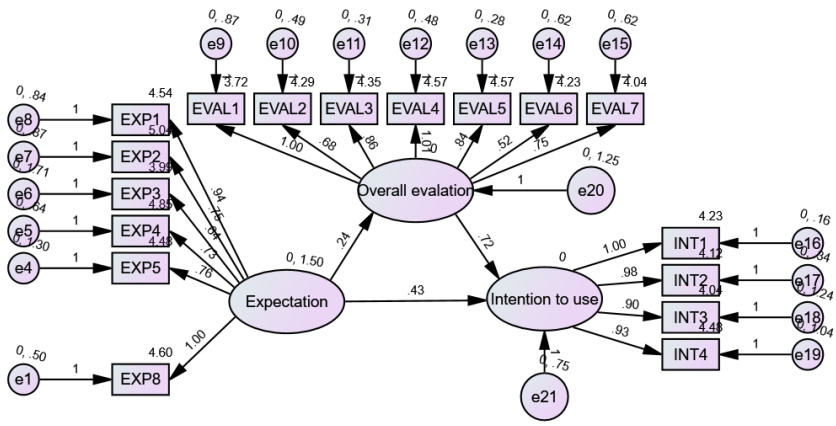


Figure 7.2.4 Path analysis: Modified model of iTV UX

Figure 7.2.4 shows the validity of each measurement variable and the results of path analysis between expectation, overall evaluation and intention to use. As can be seen,

there is a significant causal relationship between overall evaluation, and intention to use.

7.2.2 Item parceling of interactive TV UX attributes

As mentioned above, the number of data input to this model is 82. This number is usually considered very small compared to the claim that 200 or more is stable for the structural equation model. However, gathering such a large number of input data is very difficult in a lab study that measures actual user experience. In this regard, it is argued that it is reasonable to estimate the appropriate samples based on the number of free parameters that need to be estimated for the model and not based on the criteria that the number of data points should be more than 200. According to Bentler and Chou (1987), it is suitable to calculate the appropriate number of samples because the use of too many measurement variables can increase the measurement error. Based on this argument, the recommended optimal number of data in a structural equation model is 5 to 10 times the number of free parameters. Based on the reference, a structural equation model was constructed by reducing the number of items by parceling the measurement variables. Item parceling is a method to enhance the model fit by 'parcelling' - procedure for reducing the number of items by computing sums or average scores across multiple items. According to Yang et al. (2013) as well as other related researches, item parceling may remove some difficulties in fitting normality assumptions (e.g., Bandalos, 2002; Little et al., 2002; Matsunaga, 2008) and reduce model complexity, particularity when the sample size is relatively small. The sum or average scores are then used as indicators of latent factors in the SEM analysis (Little et al., 2002; Yang, Nay, & Hoyle, 2010). Parceling is creating subsets of items to be able to get the model to run and to balance out the number of indicators on each latent. Items are first combined in 'parcels' through

summing scores over item sub-groups. Contrary, many studies point out the disadvantages of item parceling. When the dimensionality of the items is not clear, parceling should not be used (Little et al., 2002). Some researchers demonstrated that using parceling increases the estimation bias (Hall, Snell, & Foust, 1999; Stephenson, & Holbert, 2003). If the number of items is too small, the probability of error increases in the estimation process (Kline, 2011). In contrast, if the number of measurement variables is increased, the measurement error may increase or model fit may be low. It is generally known that three items are appropriate for each latent variable (Little, 2013; Matsunaga, 2008). For the item parceling method, the factorial algorithm method that binds variables based on the factor loading size of the variable is applied. (Landis et al., 2000; Rogers & Schmitt, 2004; Little et al., 2013). Factorial algorithm is a parceling method for unidimensional measures that decomposes components and combines them within different parcels again.

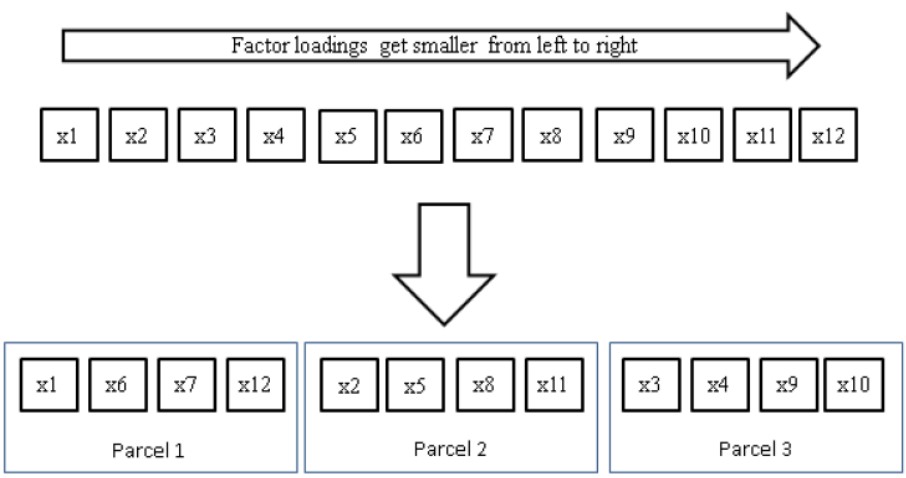


Figure 7.2.5 Factorial algorithm (Little et al., 2013, reprinted)

In summary, the original model was reviewed for regression weights, standardized regression weights, and squared multiple correlations to construct a modified model, and then the parceling of the items was repeated. The remaining variables were parceled between the largest and smallest variables according to the magnitude of factor loadings. Including bundles of items, the free parameters were initially reduced from 60 to the 17. Figure 7.2.6 below shows the result of SEM analysis after this items reducing process.

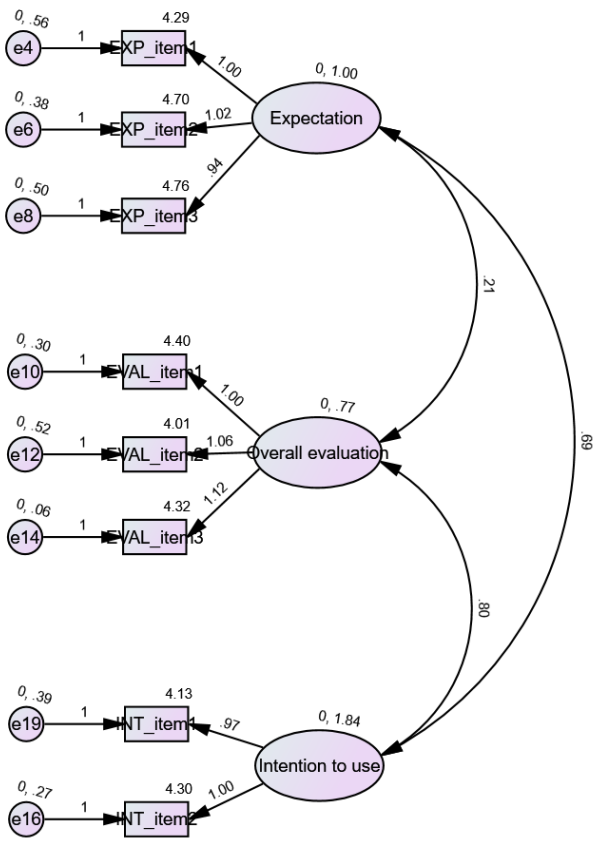


Figure 7.2.6 Modified model of interactive TV UX after item parceling

The numbers of observed variables were reduced to eight. There are three parameters for expectation, three parameters for overall evaluation, and two parameters for intention to use.

Table 7.2.5. Assignment of items to parcels

Items*	Parcel 1	Parcel 2	Parcel 3
Expectation	8th, 3rd	1st, 4th	2nd, 5th
Overall experience	4th, 6th	1st, 2nd	3rd, 5th, 7th
Continuance intention	1st, 3rd	2nd, 4th	

*Values indicate the order of factor loadings

Table 7.2.6. The results of regression analysis after item parceling

			Estimate		S.E.	C.R.
			B	β		
Overall Evaluation	<---	Expectation	0.215	0.245	0.108	1.994*
Intention to use	<---	Overall Evaluation	0.903	0.585	0.152	5.957***
Intention to use	<---	Expectation	0.494	0.364	0.133	3.722***
EXP_item1	<---	Expectation	1	0.801		
EVAL_item1	<---	Overall Evaluation	1	0.849		
EVAL_item2	<---	Overall Evaluation	1.059	0.789	0.121	8.779***
EVAL_item3	<---	Overall Evaluation	1.117	0.971	0.098	11.427***
EXP_item3	<---	Expectation	0.937	0.798	0.127	7.364***
EXP_item2	<---	Expectation	1.023	0.856	0.133	7.711***
INT_item1	<---	Intention to use	0.97	0.903	0.089	10.847***
INT_item2	<---	Intention to use	1	0.933		

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

Figure 7.2.7 is about the result of path analysis after item parceling. As shown below, the effect of overall evaluation on intention to use is slightly stronger than before reducing variables. The correlation between expectation and intention to use was same. The correlation between expectation and overall evaluation showed a decreased relation than before items reduction.

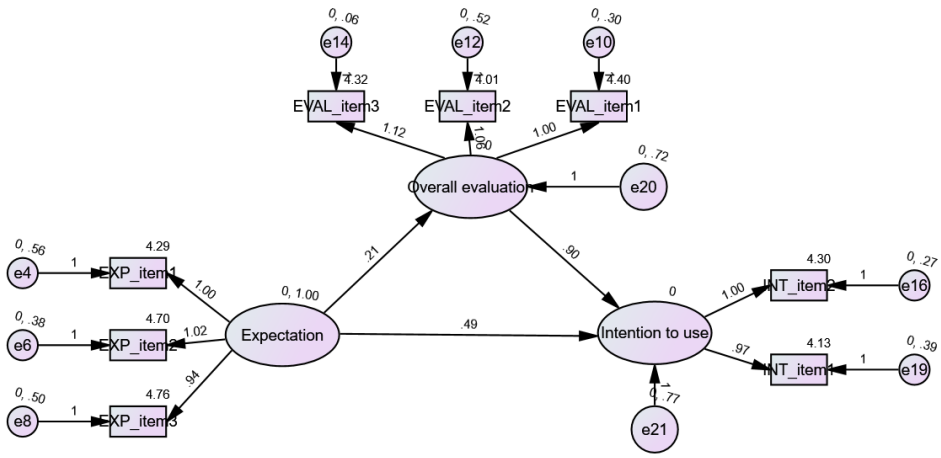


Figure 7.2.7 the result of path analysis after item parceling

Table 7.2.7 Fitness index of measurement model (after item parceling)

Model	χ^2	df	TLI	CFI	RMSEA
	40.915**	17	0.912	0.947	0.132

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

7.2.3 Confirmatory factor analysis for UX sub dimensions

Next, confirmatory factor analysis was conducted to show whether the observed

variables of pragmatic quality and hedonic quality (identification, and stimulation) are appropriate. Figure 7.2.8 shows a structural equation model for the causal relationship between the UX sub qualities—pragmatic quality, hedonic qualities (identification and stimulation), overall evaluation, and intention to use. As shown, hedonic quality is classified into two types, hedonic quality identification and hedonic quality stimulation.

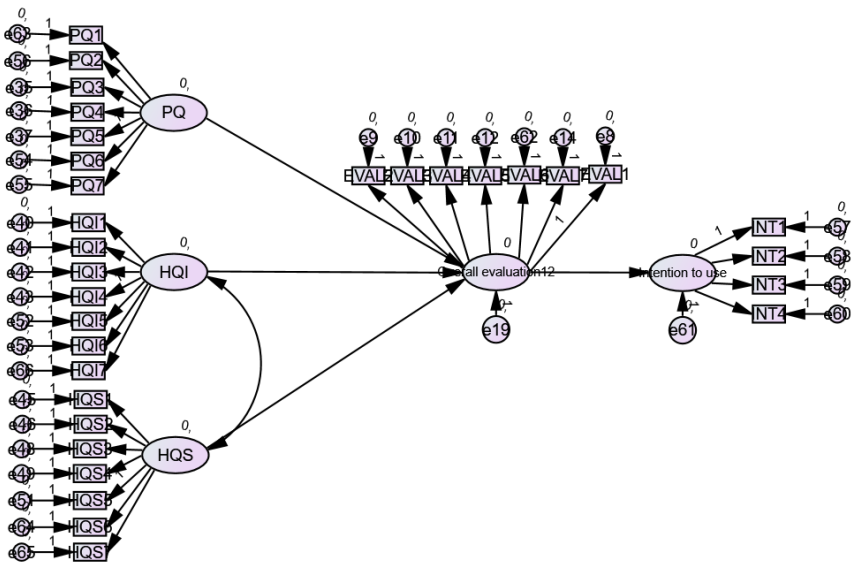


Figure 7.2.8 Analysis result of structural equation model (original model)

Table 7.2.8 Regression analysis results (original model)

			Estimate			
			B	β	S.E.	C.R.
Overall Evaluation	<---	PQ	0.408	0.451	0.079	5.162***
Overall Evaluation	<---	HQI	0.504	0.382	0.18	2.798**
Overall Evaluation	<---	HQS	0.399	0.409	0.127	3.131**

Intention to use	<---	Overall Evaluation	0.822	0.698	0.124	6.637***
EVAL_2	<---	Overall Evaluation	0.68	0.768	0.087	7.789***
EVAL_3	<---	Overall Evaluation	0.849	0.885	0.089	9.505***
EVAL_4	<---	Overall Evaluation	0.969	0.852	0.108	8.994***
EVAL_5	<---	Overall Evaluation	0.798	0.86	0.088	9.107***
EVAL_7	<---	Overall Evaluation	0.715	0.726	0.099	7.232***
HQI_4	<---	HQI	1	0.692		
HQI_3	<---	HQI	1.051	0.745	0.172	6.113***
HQI_2	<---	HQI	0.961	0.662	0.175	5.482***
HQI_1	<---	HQI	0.91	0.575	0.19	4.8***
PQ_5	<---	PQ	1	0.788		
PQ_3	<---	PQ	0.84	0.633	0.142	5.903***
HQS_4	<---	HQS	1.105	0.866	0.117	9.473***
HQS_3	<---	HQS	0.394	0.372	0.117	3.356***
HQS_2	<---	HQS	1.159	0.877	0.12	9.673***
HQS_1	<---	HQS	1.299	0.838	0.144	9.025***
HQS_5	<---	HQS	1	0.821		
PQ_4	<---	PQ	0.921	0.69	0.141	6.534***
HQI_5	<---	HQI	0.99	0.765	0.158	6.264***
HQI_6	<---	HQI	1.111	0.787	0.173	6.426***
PQ_6	<---	PQ	1.024	0.916	0.111	9.209***
PQ_7	<---	PQ	1.025	0.817	0.127	8.061***
INT_1	<---	Intention to use	1	0.967		
INT_2	<---	Intention to use	0.965	0.912	0.062	15.511***
INT_3	<---	Intention to use	0.904	0.757	0.095	9.537***
INT_4	<---	Intention to use	0.919	0.784	0.089	10.28***
EVAL_6	<---	Overall Evaluation	0.489	0.589	0.088	5.586***
EVAL_1	<---	Overall Evaluation	1	0.802		
PQ_2	<---	PQ	0.644	0.521	0.136	4.732***

PQ_1	<---	PQ	-0.009	-0.007	0.145	-0.06
HQS_6	<---	HQS	0.757	0.602	0.131	5.791***
HQS_7	<---	HQS	0.935	0.772	0.117	7.999***
HQI_7	<---	HQI	1.143	0.672	0.205	5.56***

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

As a result of the original model measurement, the model fit does not reach the goodness-of-fit (χ^2 = 493.568 (p < .001), TLI = .698, CFI = .732, RMSEA = .143).

Table 7.2.9 Fitness index of measurement model (original)

Model	χ^2	df	TLI	CFI	RMSEA
	970.678***	62	0.733	0.753	0.117

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

In order to improve the model fit in the initial model, “cautious – bold” (HQS3), one of the attributes of Hedonic Quality Stimulation and “technical – human” (PQ1) which is a parameter of Pragmatic Quality were removed. Next, the model fit was examined by removing variables that do not satisfy Estimate> 0.4 from Squared Multiple Correlations. “Complicated – simple” (PQ2), “isolating – connective” (HQI1), “impractical – practical” (PQ3), “unpresentable – presentable” (HQI7), “cumbersome-straight forward” (PQ4) and “unprofessional – professional” (HQI2) were removed one by one (Table 7.2.10).

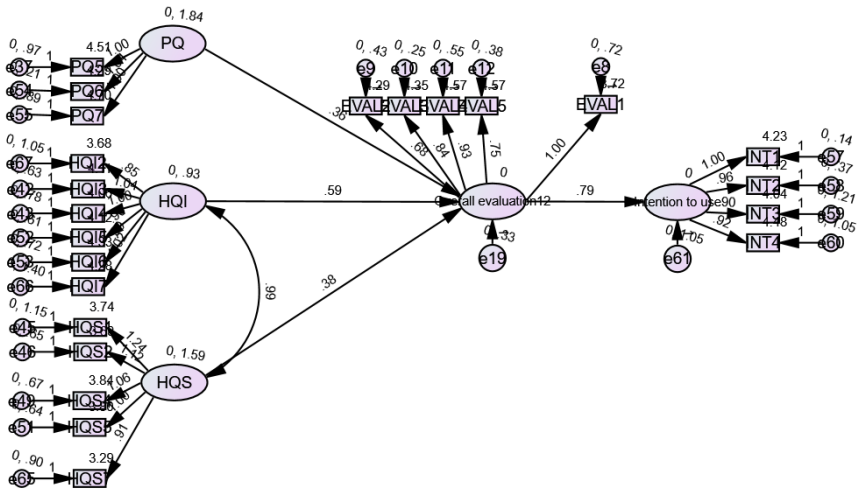


Figure 7.2.9 Analysis result of structural equation model of UX factors (modified)

Table 7.2.10 Regression analysis results (modified model)

			Estimate			
			B	β	S.E.	C.R.
PQ7_manageable	<---	Pragmatic Quality	1	0.829		
PQ6_ clearly structured	<---	Pragmatic Quality	0.998	0.929	0.102	9.805***
PQ5_ predictable	<---	Pragmatic Quality	1.014	0.831	0.115	8.837***
HQI3_ stylish	<---	Hedonic Quality Identification	0.986	0.832	0.116	8.471***
HQI4_ premium	<---	Hedonic Quality Identification	1	0.823		
HQI5_ integrating	<---	Hedonic Quality Identification	0.728	0.67	0.114	6.396***
HQI6_ bring me closer	<---	Hedonic Quality Identification	0.779	0.657	0.125	6.245***
HQS1_ inventive	<---	Hedonic Quality	1	0.863		

		Stimulation				
HQS2_innovative	<---	Hedonic Quality Stimulation	0.85	0.86	0.083	10.262***
HQS4_ innovative	<---	Hedonic Quality Stimulation	0.825	0.864	0.08	10.341***
HQS5_ captivating	<---	Hedonic Quality Stimulation	0.709	0.777	0.082	8.624***
HQS6_ challenging	<---	Hedonic Quality Stimulation	0.614	0.653	0.092	6.661***
HQS7_ novel	<---	Hedonic Quality Stimulation	0.726	0.8	0.08	9.05***

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

After removing these variables, the model fit improved to $\chi 2$ = 129.054 (p <.001), TLI = .885, CFI = .909, and RMSEA = .16 (see Table 7.2.11). A possible explanation for this is that the usability qualities of different interfaces (CA and RCU) are opposite to each other.

Table 7.2.11 Fitness index of measurement model (modified)

Model	χ^2	df	TLI	CFI	RMSEA
	129.054***	62	0.885	0.909	0.16

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

Overall evaluation and hedonic quality identification showed stronger causality than pragmatic quality. Pragmatic quality was weak, but causality was associated with intention to use, while hedonic quality identification was causal only with the overall experience. Intention to use and hedonic quality did not appear to be causal, suggesting that HQ does not directly affect intention to use.

7.2.4 Item parceling of UX sub dimensions

The Figure 7.2.10 shows the original model that includes both the process by which the user experience is formed and the sub factors of the user experience. However, this model was judged to be unsuitable as structural equation model because there are 70 measurement variables to data number. First, the variables that did not meet the criteria were removed. Next, the latent variables with similar factor loadings of the sub-measured variables were combined with the sub-measured variables and converted into measurement variables.

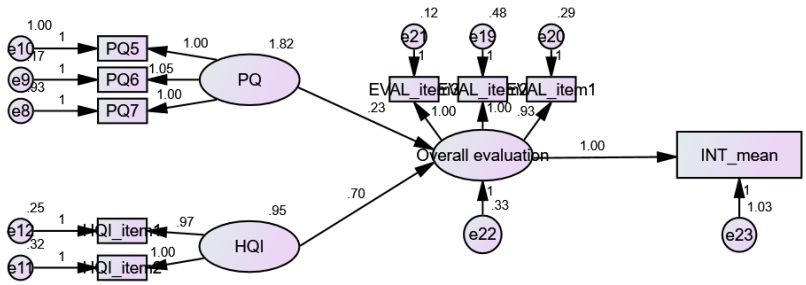


Figure 7.2.10 Structural equation model of UX attributes (modified and item parceled)

The results of regression analysis for UX attributes after item parceling is shown below (Table 7.2.12).

Table 7.2.12 Regression analysis results after item parceling of UX attributes

			Estimate B	S.E.	C.R.
Overall Evaluation	<---	PQ	0.23	0.063	3.636***

Overall Evaluation	<---	HQI	0.7	0.111	6.287***
PQ_7	<---	PQ	1		
PQ_6	<---	PQ	1.052	0.109	9.637***
PQ_5	<---	PQ	0.997	0.119	8.396***
HQI_item2	<---	HQI	1		
HQI_item1	<---	HQI	0.967	0.115	8.436***
EVAL_item2	<---	Overall Evaluation	1		
EVAL_item1	<---	Overall Evaluation	0.928	0.103	8.968***
EVAL_item3	<---	Overall Evaluation	0.998	0.099	10.127***

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

The numbers of free parameters were down to 18 that the required subject numbers were satisfied. After parceling items, the model fit slightly improved to $\chi 2 = 52.109$ ($p < .001$), TLI = .917, CFI = .943. However, the result of RMSEA was slightly increased (see Table 7.2.13).

Table 7.2.13 Fitness index of measurement model (after item parceling)

Model	χ^2	df	TLI	CFI	RMSEA
	54.109**	18	0.917	0.943	0.120

*indicates level of significance. *** *p* value < .001, ** *p* value < .01, * *p* value < .05

7.3 Summary

Unlike the interface comparison research in the previous chapter, this chapter examined the validity and causality of the measurement variables to construct a general interactive TV UX framework. The results confirmed that there is a static causal relationship between expectation and overall user evaluation and intention to use. It was also found that expectation, pragmatic quality, hedonic quality identification, and hedonic quality stimulation categories are suitable parameters for

use as UX measurement measures. However, fun and rewarding of the expectation category were found to be unsuitable for constructing the structural equation model. This was eliminated for the same reason in the path analysis. Therefore, it seems reasonable to avoid selecting those measurement variables in interactive TV UX studies in the future. In detailed route analysis, hedonic quality stimulation was eliminated as a latent variable since it showed no causal relationship to both, overall evaluation or the intention to use. This result implies that hedonic quality stimulation depends on the difference in the user interface, as shown in the previous study. In summarizing, users in RCU and CA experienced the biggest difference in terms of hedonic quality stimulation, which indicates that it is difficult to generalize the UX research scale. It also means that the applicable variables can be different in measuring user experience between different interfaces. This suggests that selecting appropriate variables according to the interface is also necessary to obtain meaningful data. The results of the hypothesis testing using the structural equation model are summarized as follows.

Table 7.3.1 Summary of hypothesis tests

Hypotheses	B	β	C.R.	Result
H1: behavioral intentions ← overall evaluation	0.688	0.587	5.944	Approved
H2: behavioral intentions ← perceived pragmatic quality				Rejected
H3: behavioral intentions ← perceived hedonic quality identification				Rejected
H4: behavioral intentions → expected satisfaction	1.325	0.309	4.284	Approved
H5: overall evaluation → expected satisfaction				Approved
H6: overall evaluation → perceived pragmatic quality	0.211	0.244	2.956	Approved
H7: overall evaluation → perceived hedonic quality identification	0.899	0.804	5.597	Approved

Both the perceived pragmatic quality and the perceived hedonic quality during UX phase had a significant effect on the overall user evaluation, respectively. Overall

user evaluation indicates overall user satisfaction. Overall user experience was found to have a significant effect on both the intention to use of CA and RCU. Pragmatic quality and hedonic quality identification showed significant correlations among expected user satisfaction, overall user experience, and intention to use. Hedonic quality stimulation, however, did not affect experienced satisfaction or the intention to use. Due to a lack of subjects, the results of long-term UX study of the last phase of the interactive TV UX model framework was not included in the analysis of the structural equation models.

Table 7.3.2 UX attributes affecting interactive TV

	Study 1	Study 2 (RCU)	Study 2 (CA)	SEM
Pragmatic Quality	Simple **Controllable**	**Manageable** Clearly structured predictable		**Manageable** Clearly structured Predictable
Hedonic Quality	**Innovative** **Impressive**		Stylish Premium Integrating Affinity Inventive **Innovative** **Captivating** Challenging	Stylish Premium Integrating Affinity Inventive **Innovative** **Captivating** Challenging
Overall user evaluation	**Original** Good Motivating		**Novel** Pleasant Attractive Likeable Inviting	**Novel** Good Motivating Attractive Likeable Inviting Appealing

Table 7.3.2 shows the attributes derived from SEM and those results from the previous two studies. The measured variables derived from the structural equation

model for iTV UX are “manageable,” “clearly structured,” “predictable,” “premium,” “innovative.” Compared with the results of previous two experiments, “manageable,” “innovative,” “captivating,” and “novel” are considered to be the common attributes contributing to a positive iTV user experience. “Manageable” is one of attributes in the pragmatic quality dimension. “Innovative” and “captivating” are the attributes of hedonic quality. The quality attribute influencing to the overall user evaluation was “novel.” The main attributes affecting the quality of the overall user experience presented above are described in the diagram as follows (Figure 7.3.1). The figure below shows the process of user experience and the main factors that affect the user experience at each phase.

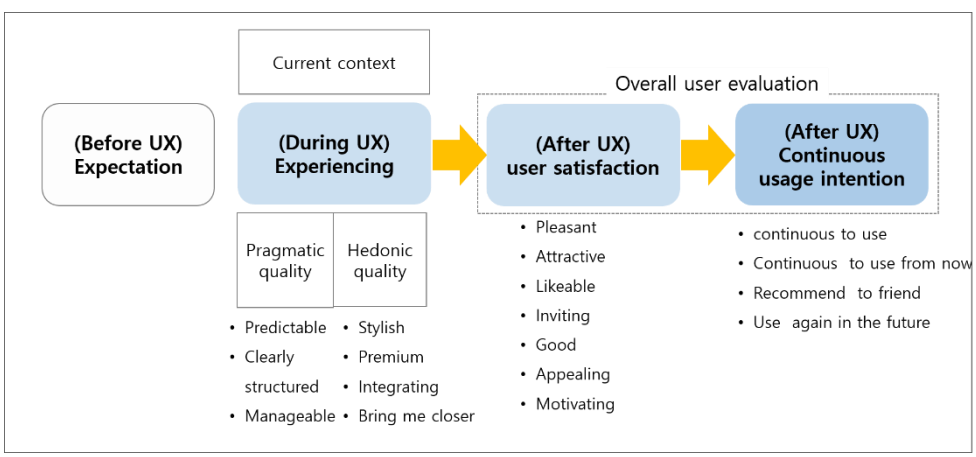


Figure 7.3.1 Modified interactive TV UX framework with affective UX attributes

In this study, iTV UX model framework was developed based on two previous studies: Mahlke and Thüring 's (2007) UX model framework and Hassenzahl’s (2003) UX measurement parameters. The UX model presented by these previous studies gives a very detailed description of the process of the UX sub-dimensions and the

overall evaluation However, the properties of overall evaluation are presented in the same dimension without describing the relationship between the properties. If user experience studies are to be applied to actual technical strategies or design direction, there is evidence that they are linked to more specific steps than user satisfaction, that is, to the user's behavioural intention to use. In this study, it was identified the causal relationship between the UX factors and user satisfaction, and user satisfaction and the behavioural intention, respectively. This supports the more tangible "user satisfaction" and exploring user experience is a practical study contributing industrially as well as academically.

The advent of UX has drawn attention to the importance of emotional quality, but t hese properties have characteristics that are difficult to concretely prove. It was als o constrained in the field of HCI, which requires more objective and quantitative da ta. The technology acceptance model can be a system perspective evaluation or an early technology forecast data, but this alone is very limited as research data for ext ending the user experience. In this study, it was confirmed that the actual user expe rience has a relationship to the intention to use, and user satisfaction is the key to li nk each side together. The influence of expectation without "experiencing" was ver y small. Satisfaction through the specific use experience predicted the continuous u se of the technology.

In addition, some of the measurement parameters presented in Hassenzahl's UX model were eliminated in structural equation model analysis. This also means that for each product, measurement parameters that appeal to the experience of use are necessary to be adjusted. The hedonic quality that the user has in the expectation phase is not related to the UX qualities created after the experience. In other words, emotional enjoyment is based on actual user experience. The variables such as fun

and rewarding in the before UX phase were significantly different from the experience qualities in the after UX phase. The reason why HQS showed no causal relationship with overall satisfaction and intention to use is that the stimulation quality between RCU and CA is opposite to each other. As a result, the sum of the results between the two interfaces could not be used as significant data. In other words, PQ and HQI are considered to have a significant influence on overall satisfaction despite the difference in interface. However, HQS is the opposite between interfaces, and this result is not used as a variable, indicating a positive causal relationship to overall satisfaction. In conclusion, HQS exhibits different quality depending on the interface, and is not suitable for usage as a latent variable for overall experience quality.

Chapter 8
Discussion and conclusion

8.1 Discussion

This study begins with the definition of the experience of using interactive products and derives the measurement scale through reviews of existing studies to measure the quality of experience. Because UX ranges are broad and ambiguous, it is necessary to clarify the ranges and predetermined methods, objectives, theories, and so on. Next, interactive TV UX model and related attributes of UX quality dimensions from previous studies were reviewed. The summary of the study is as follows:

The first purpose of this study is to find the UX quality attributes affecting user satisfaction based on the user's perspective. For this purpose, two empirical studies were conducted considering the user context and the user interface. In evaluating the user experience according to the different behavioral situations, it was confirmed that users are more appealed in the goal-directed situation. The observed variables that contributed to the positive user experience were "simple" and "controllable" in the pragmatic quality dimension, and "innovative" and "impressive" in the hedonic quality dimension. In the user interface evaluation, a comparison evaluation for operating TV using RCU and CA was performed. In RCU interaction, "manageable," "clearly structured" and "predictable" were found that affect user satisfaction. CA,

on the other hand, had more impact on TV viewer's hedonic quality. "Premium," "integrating," "innovative," "captivating" and "inventive" are affecting variables in hedonic quality dimension. Innovative and captivating are same variables which are found in the study of user behavioral situations. The measured variables derived from the structural equation model for iTV UX are "manageable," "clearly structured," "predictable," "premium," "innovative." Compared with the results of previous two experiments, "manageable," "innovative," "captivating," and "novel" are considered to be the common attributes contributing to a positive iTV user experience. "Manageable" is one of attributes in the pragmatic quality dimension. "Innovative" and "captivating" are the attributes of hedonic quality. The quality attribute influencing to the overall user evaluation was "novel." In all experiments, HQ had more impact than PQ on the positive user experience of interactive TV. In summary, the attributes derived from this study re the key variables contributing to the positive experience of interactive TV. TV viewers want to freely manage in interacting with TV, and think more valuable when they feel innovative and captivating in the TV interaction. Also, "novel" may be seen as a discriminatory attribute that distinguishes interactive TV from the traditional TV.

Second, it was confirmed that CA plays a role of complementary rather than replacing RCU as an interface of interactive TV, and it was also found that CA can lead a more positive experience in emotional aspects. RCU is suitable for basic TV operation, and CA can be used as a tool for the extended experience such as recommending content and providing usable information. In worldwide, CA is competitively distributing because it is expected to become the next platform of home appliances. However, in Korea, it is still under development and CA equipped devices are communicated in the technology perspectives rather than user oriented experience. Design strategies and scenario development using CA is necessary to

extend usage of CA, and these needs require UX based researches for every day products. As expected, the usability quality of the CA was still low, but the emotional quality evaluation showed that the CA was much higher than the RCU. This quality also had a positive effect on user satisfaction and willingness to continue use. By applying a variety of methods for UX measurement, it broadened to understand the advantages and disadvantages of each method applicable to the UX process. In particular, the measurement of emotional response is highly necessary, since it can identify immediate user response during use, not post-use.

As a result of the biological response, the user was able to see that tension was lower before the start of using the CA. This difference in response time is very useful in that it is possible to identify the specific cause of where the positive or negative user experience comes from. Finally, it is confirmed whether the user experience affects the willingness to adopt new technologies. To this end, a user experience study on a CA-equipped TV was conducted. Although CAs are commercially available in many forms, complaints about overall quality of use are still high. However, it is necessary to understand whether this inconvenience is technical or emotional. If it is technical dissatisfaction, it can be overcome through improvement, and if it does not touch or cause irritation in the emotional aspect, it is likely to be ignored by the user. However, what I found through this experiment is that CA and RCU can complement each other. Currently, TV interfaces are provided around RCU-based operation scenarios, but in the future, development of operation scenarios involving CAs is also required.

TV interfaces are designed according to RCU-based operation scenarios. In the future, development of operation scenarios involving CA may be required. Luger and Sellen (2016) interviewed 14 CA users and found that there was still a gap between the expectations and experience of using CA. In particular, unlike expert

users, novice users who are unfamiliar with CA had higher expectations for the system in before usage, and were more embarrassed by the results. In many cases, failure to communicate or fail to achieve the desired purpose is often abandoned or not used. Currently, the CA function has been added without changing the existing TV viewing context. However, in the future, it is expected that the interface design and the scenario for utilizing the functional aspect of the CA and its emotional properties will change. It is expected that the TV viewing context will be changed and that the TV experience will be strengthened. Lastly, the interactive TV UX measurement model is constructed and verified.

Third, an interactive TV UX model framework derived from a phase based research process was presented and verified. The relationship between major UX factors is constructed by structural equation model. This ensured that user experience quality contributes not directly to behavioural usage intention, but to the user satisfaction, and when users are satisficed through the interaction, then the user satisfaction affects the intention to use. In this study, iTV UX model framework was developed based on two previous studies: Mahlke and Thüring's (2007) UX model framework and Hassenzahl's (2003) UX measurement parameters. The UX model presented by these previous studies gives a very detailed description of the process of the UX sub-dimensions and the overall evaluation However, the properties of overall evaluation are presented in the same dimension without describing the relationship between the properties. If user experience studies are to be applied to actual technical strategies or design direction, there is evidence that they are linked to more specific steps than user satisfaction, that is, to the user’s behavioural intention to use. In this study, it was identified the causal relationship between the UX factors and user satisfaction, and user satisfaction and the behavioural intention, respectively. This supports the more tangible “user satisfaction” and exploring user experience is a practical study

contributing industrially as well as academically. The advent of UX has drawn attention to the importance of emotional quality, but these properties have characteristics that are difficult to concretely prove. It was also constrained in the field of HCI, which requires more objective and quantitative data. The technology acceptance model can be a system perspective evaluation or an early technology forecast data, but this alone is very limited as research data for extending the user experience. In this study, it was confirmed that the actual user experience has a relationship to the intention to use, and user satisfaction is the key to link each side together. The influence of expectation without “experiencing” was very small. Satisfaction through the specific use experience predicted the continuous use of the technology. In addition, some of the measurement parameters presented in Hassenzahl's UX model were eliminated in structural equation model analysis. This also means that for each product, measurement parameters that appeal to the experience of use are necessary to be adjusted.

8.2 Contribution of this study

The significance of this study is as follows. First, iTV UX research was conducted from the user's perspective, considering various user context and environment, not from the technology perspective. Many interactive TV related researches have focused on hardware aspects such as EPG, on demand video content, network, and quality of picture and sound. These research trends are required to change as TV evolve into interactive products. The reason why users do not use the new features of interactive TV actively is that the environment of interaction between the system and the user is not optimized yet. The interface in the interactive TV is similar to that of smartphone and PC. RCU, the main controller of TV is not suitable for the

interactive environments such as searching and recommending content. The iTV user environment should be designed to be suitable for watching TV with relaxation and fun, and this design strategy should start with what experience will provide to the users. Experience-based research is an initial and basic approach in which a TV user has an experience, and therefore, a user's experience is more diffused and enhanced only by stimulating or reinforcing a certain quality of experience. This is because any product or technology to be designed for "usage" is alive when the user “use” it. And the term “usage” indicates that it has emotional appeal as well as usable aspects.

Second, it was verified that integrated analysis for the overall user experience is necessary through the various application of multiple methods, breaking away from the constraints of adopting the single method (Maja and Furtado, 2016). In this study, various aspects of interactive TV UX were considered and evaluated, according to the behavioural situations, user interfaces and prolonged use. Each study used mixed methods such as physiological measurements, user self-report questionnaires, and user diaries and interviews. Physiological measurements were performed to see the emotional response during interaction. The subjective method of evaluating the user's own experience after the interaction was also applied. As a result, it was also confirmed there are correlations between the physiological recordings, an objective method, and users self-report questionnaires, a subjective method. This result leaded that applying the combined methods for evaluating user experience are suitable process to improve the validity and the reliability of UX studies.

Field study for 20 households for four weeks was also performed to measure long-term user experience in daily lives. The results showed that user satisfaction increased over time, and there was a significant correlation between the results of perceived user satisfaction and the continuous usage intention. The comparison

between the user data of short term study and that of long term field test showed the similar results. Similar to the result of short-term experiment, field study showed that the positive user experience has a causal relationship with user satisfaction and intention to use. This implies it is necessary to consider not only short-term study, but also to see how the user satisfaction changes as a result of prolonged use. Third, this study revealed that UX study can deliver meaningful output to plan specific design strategies or predict technology acceptance, rather than phenomena studies or abstract results. Improvements of existing products can be made through customer surveys or usability assessments.

However, identifying customer satisfaction or predicting the continuous usage intention of a new product is not possible with a usability assessment alone. The intention to continue to use the new technology can only be predicted based on the user's satisfaction with the user experience. According to Sanders (2001), experience is linked to past memories and imagination of future use. UX is vital for the success of interactive products. In this study, it was constructed a user experience model centering on interactive TV and conducted a user experience study from various viewpoints by applying complex evaluation methods. It is hoped that these studies can contribute a little to creating products and services that are satisfactory and consistently available to users. Currently, user experience models that combine cognition and emotion become the common paradigm (Helander and Khalid, 2006) and it is very clear that it is important to have overall user interface including emotion and affects beyond usability. I hope that this study will contribute to providing meaningful findings in research on interactive technology and user experience mediated by conversational agents. Questions regarding behavioral intentions (Kujala et al. 2017; Reichheld, 2003; Wirtz et al., 2003) seek to obtain basic information regarding the willingness to continually use the product, willingness to

recommend it to a friend, and the intention to use the product/service in the future. These questions focus more on the attitude of the user than on the specific extent or scope of use. In other words, if simple questions that do not specifically limit the range or purpose of use are asked, users may provide answers with different criteria. In future research, questions that are more elaborate are required to obtain more specific user feedback about the products/services.

8.3 Limitation and further study

This study has the following limitations in several aspects: This study was conducted to evaluate interactive TV UX and studying CA as a TV interface and confining it to research on TV-related tasks, it was not able to access CA properties more in depth and stayed in fragmentary studies. As an alternative to the RCU, the same task was performed to evaluate the CA. However, the actual CA speech recognition error frequently appeared, resulting in a very low CA performance compared to the RCU. Given that the performance difference between the two devices is very large because the click failure rate of the RCU is nearly zero, the next study needs to present an optimized environment for each device and conduct the study rather than evaluating the devices one-to-one. It is considered. On the other hand, frequent errors in CA speech recognition are a reflection of reality, and therefore, there are many constraints on usability and usability, thus revealing the necessity of suggesting an optimized usage environment considering these limitations. In the Korean market, CA is spreading through other devices, led by telecommunication companies, which indicates that it is not user-driven consumption. Having a user use a device that has not been optimized for use can have a very negative experience with the CA. In particular, the greater the disappointment to the expectation, the more negative the

experience. Therefore, it is also necessary to have a guideline to control the CA by recognizing the level of CA quality and learning about the scope or method of use.

In this study, very limited users data in SEM analysis were conducted for the developing the interactive TV UX model framework. Compared to the general structural equation model requiring more than 200 data, there is a limitation to support the reliability of this study. Most of large-scaled user researches are conducted through online or offline surveys. This study is about the user experience, and all experiments were conducted by evaluating actual users' experience through interaction with TV. To compensate for this, the item parcelling was conducted to improve the model fit by parcelling the variables according to the magnitude of factor loadings. But even with a small number of measurement data, the results of structural equation modelling for iTV confirmed that the overall user experience acquired through UX sub-dimensions (PQ and HQ) is directly related to the user satisfaction, and it also affects the intention to use new technology. Expectation had only a limited effect to the user's behavioural intention. This implies that research on real user experience is more important for predicting the viability of new technology and only expectation are not even though the study is about user experience. It also implies that judging whether or not to accept technology based on expectation alone may not be a suitable user experience study.

Further study is necessary on the conversational agent. The voice recognition function will be combined indispensably with various products around us in the near future. This is because it is not just "smart" but also a method that "stimulates emotion" and more natural interaction. In this aspect, it will be a meaningful attempt to use the physiological measurements to directly determine the physical responses occurring immediately in the interaction process. People think that they know their

emotions or feelings and express them properly, but in reality, it is not true. Particularly, it is almost impossible to remember completely or determine accurately the instantaneous changes of emotion or the feelings that change every moment. Natural interaction requires a physical response measurement, and through this, full explanation can be provided on how the UXs were constructed. Nevertheless, for the part regarding which result will be obtained by applying which method, diverse studies are required according to the target situation. The technological evolution of the CA is ongoing and at the same time, various perspectives are being considered.

In the 2016 IBM Developer Day, Michal Yuan (2016) emphasized two things for success factors when designing a CA system: the CA has to be operated in a limited scope, and by providing characteristics suitable for the usage scope and goal of the CA, a context-model design for conversation with the user should be properly done. Other parts that should be considered along with the technological evolution of the CA are non-linguistic expressions such as sighs, hand gestures, and facial expressions. These are major signals that are referenced when I communicate with a counterpart. In order to infuse the CA naturally into the daily lives of users, studies on various non-linguistic expressions (e.g., tones and nuances) should be carried out in parallel with efforts to improve the recognition rate of voice. If personified cues (anthropomorphic or personification cue) are added, more natural and ordinary interactions will be possible. Through such a process, the voice recognition technology will be combined with our everyday products as an “intelligent agent” instead of just being “smart” interaction tools. Although a review on the effective implementation method of the CA is outside the scope of this study, continuous research on this is required. Reeves (2017) stated that these devices present challenging tasks in the computer-supported cooperative work and HCI areas because they throw fundamental questions regarding “cooperative interaction design

between heterogeneous human-agent groups" and "interaction between humans and machines."

8.4 Conclusion

So far, it has been described that the research on interactive TV UX, based on theoretical inquiries, process arrangement, and the results of empirical studies on the TV. TV is one of the most familiar products in our lives, and it is likely one of the products to stay with us for a long time. TV is evolving through new technologies and changing our experience according to the evolution. In the HCI area, the usability perspective that has been accumulated for a long time has extended to the UX framework as the importance of emotional quality has risen. On top of this, the active adoption of natural interactions of humans such as hearing and tactile sense, which go beyond the visual interaction, requires UX studies in more multidisciplinary and in-depth perspective. Continuing to observe and explore the changes of everyday products which are accustomed to us is the most essential to enrich our daily lives. It is hoped that this study will be applied to the basic study of user experience of the everyday products such as TV, so that users can have a more positive and enriched experience.

Interactive TV experience studies can also be applied to assess the experience of other everyday products evolving into interactive systems. The evolution of everyday products is now becoming phenomenon. Studying user experiences in interactive products could help not only new technologies, but also interface strategies for all everyday products that evolve into interactive products. This paper focuses to discuss how users can embrace and experience existing products with new technologies, and how they accept new technologies by "experiencing." Although

this study does not cover everything about iTV UX, I believe it would be a worthwhile if this research could give any meaningful results to the HCI field.

References

Amin, M., Rezaei, S. and Abolghasemi, M., 2014. User satisfaction with mobile websites: the impact of perceived usefulness (PU), perceived ease of use (PEOU) and trust. Nankai Business Review International, 5(3), pp.258-274.

Apter, M.J., 1989. Reversal theory: Motivation, emotion and personality. Taylor & Frances/Routledge.

Aranyi, G. and Van Schaik, P., 2015. Modeling user experience with news websites. Journal of the Association for Information Science and Technology, 66(12), pp.2471-2493.

Ardito, C., Buono, P., Costabile, M.F., De Angeli, A. and Lanzilotti, R., 2008. Combining quantitative and qualitative data for measuring user experience of an educational game. In International Workshop on (Vol. 27).

Babin, B.J. and Babin, L., 2001. Seeking something different? A model of schema typicality, consumer affect, purchase intentions and perceived shopping value. Journal of Business research, 54(2), pp.89-96.

Bandalos, D.L., 2002. The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. Structural equation modeling, 9(1), pp.78-102.

Bargas-Avila, J.A. and Hornbæk, K., 2011, May. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 2689-2698). ACM.

Basapur, S., Harboe, G., Mandalia, H., Novak, A., Vuong, V. and Metcalf, C., 2011, June. Field trial of a dual device user experience for iTV. In Proceedings of the 9th European Conference on Interactive TV and Video (pp. 127-136). ACM.

Batra, R. and Ahtola, O.T., 1991. Measuring the hedonic and utilitarian sources of consumer attitudes. Marketing letters, 2(2), pp.159-170.

Becker, C., Kopp, S. and Wachsmuth, I., 2007. Why emotions should be integrated into conversational agents. Conversational informatics: an engineering approach, pp.49-68.

Bellman, S., Robinson, J.A., Wooley, B. and Varan, D., 2017. The effects of social TV on television advertising effectiveness. Journal of Marketing Communications, 23(1), pp.73-91.

Bengs, A., Wiklund-Engblom, A., & Sperring, S. A. 2009. Multimethod Approach for Measuring Mobile UX. In MobileHCI (Vol. 9, pp. 15-18).

Bentler, P.M. and Chou, C.P., 1987. Practical issues in structural modeling. Sociological Methods & Research, 16(1), pp.78-117.

Belanche, D., Casaló, L.V. and Guinalíu, M., 2012. Website usability, consumer satisfaction and the intention to use a website: the moderating effect of perceived risk. Journal of retailing and consumer services, 19(1), pp.124-132.

Bernhaupt, R. & Pirker, M. (2013, September). Evaluating user experience for interactive television: towards the development of a domain-specific user experience questionnaire. In *IFIP Conference on Human-Computer Interaction* (pp. 642-659). Springer, Berlin, Heidelberg.

Beyah, G., Xu, P., Woo, H., Mohan, K. and Straub, D., 2003. Development of an instrument to study the use of recommendation systems. AMCIS 2003 Proceedings, p.34.

Bhattacherjee, A. and Premkumar, G., 2004. Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. MIS quarterly, pp.229-254.

Blaynee, J., Kreps, D.D., Kutar, D.M. and Griffiths, D.M., 2016, July. Collaborative HCI and UX: longitudinal diary studies as a means of uncovering barriers to digital adoption. In Proceedings of the 30th

International BCS Human Computer Interaction Conference: Fusion! (p. 53). BCS Learning & Development Ltd.

Blythe, M., Hassenzahl, M., Law, E. and Vermeeren, A.P.O.S., 2007. An analysis framework for user experience (UX) studies: A green paper. Towards a UX Manifesto, 1(140), p.6.

Boomsma, J.J. and Van Loon, A.J., 1982. Structure and diversity of ant communities in successive coastal dune valleys. The Journal of Animal Ecology, pp.957-974.

Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: Comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. International Journal of Human-Computer Interaction, 31(8), 484-495.

Brajnik, G. & Giachin, C. (2014). Using sketches and storyboards to assess impact of age difference in user experience. International Journal of Human-Computer Studies, 72(6), 552-566.

Bushnell, M. (2018, June 29). AI Faceoff: Siri vs. Cortana vs. Google Assistant vs. Alexa. Business News Daily. Retrieved from https://www.businessnewsdaily.com/10315-siri-cortana-google-assistant-amazon-alexa-face-off.html

Cen, L., Wu, F., Yu, Z.L. and Hu, F., 2016. A real-time speech emotion recognition system and its application in online learning. In Emotions, technology, design, and learning (pp. 27-46). Academic Press.

Cesar, P. and Chorianopoulos, K. (2008). October. Interactivity and user participation in the television lifecycle: creating, sharing, and controlling content. In Proceedings of the 1st international conference on Designing interactive user experiences for TV and video (pp. 125-128). ACM.

Chapanis, A., 1995. Ergonomics in product development: a personal view. Ergonomics, 38(8), pp.1625-1638.

Chitturi, R., Raghunathan, R. & Mahajan, V. (2008). Delight by design: The role of hedonic versus utilitarian benefits. *Journal of marketing*, 72(3), 48-63.

Chorianopoulos, K. & Spinellis, D. (2004). Affective usability evaluation for an interactive music television channel. Computers in Entertainment (CIE), 2(3), 14-14.

Chung, N. and Kwon, S.J., 2009. The effects of customers' mobile experience and technical support on the intention to use mobile banking. Cyberpsychology & Behavior, 12(5), pp.539-543.

Cheong, J. and Park, M. 2005. Mobile internet acceptance in Korea. Internet Research, 15: 125–140.

Cockton, G. (2002). From doing to being: bringing emotion into interaction, Interacting with Computers, Volume 14, Issue 2, February 2002, Pages 89–92.

Cockton, G., 2006, October. Designing worth is worth designing. In Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles (pp. 165-174). ACM.

Consumer Insight (2018). *The 27th Mobile Communication Research in Korea*. Retrieved from https://www.consumerinsight.co.kr/voc_view.aspx?no=2924&id=ins02_list&PageNo=1&schFlag=0

Damasio, A.R., Tranel, D. and Damasio, H.C., 1991. Behavior: theory and preliminary testing. Frontal lobe function and dysfunction, 217.

Darnell, M. J. (2007). How do people really interact with TV?: naturalistic observations of digital tv and digital video recorder users. Computers in Entertainment (CIE), 5(2), 10.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Quarterly, 13, 31-340.DeLone and McLean, 1992

Davis, R., Lang, B. and Gautam, N., 2013. Modeling utilitarian-hedonic dual mediation (UHDM) in the purchase and use of games. Internet Research, 23(2), pp.229-256.

DeLone, W.H. and McLean, E.R., 1992. Information systems success: The quest for the dependent variable. Information systems research, 3(1), pp.60-95.

Deng, L., Turner, D.E., Gehling, R. and Prince, B., 2010. User experience, satisfaction, and continual usage intention of IT. European Journal of Information Systems, 19(1), pp.60-75.

Desmet, P. and Hekkert, P., 2007. Framework of product experience. International journal of design, 1(1), pp.57-66.

Dillon, A., 2002. Beyond usability: process, outcome and affect in human-computer interactions.

Drucker, S. M., Glatzer, A., De Mar, S., & Wong, C. (2002, April). 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). ACM.

Drouet, D. & Bernhaupt, R. (2016). User Experience Evaluation Methods: Lessons Learned from an Interactive TV Case-Study C. Bogdan et al. (Eds.): HCSE 2016/HESSD 2016, LNCS 9856, pp. 351–358, 2016. DOI: 10.1007/978-3-319-44902-9_22.

Ekman, P. (1999). Basic emotions. T. Dalgleish, and M. Power (Eds). Handbook of Cognition and Emotion (pp. 45 – 60). Chichester, Sussex: John Wiley & Sons Ltd.

Flavián, C., Guinalíu, M. and Gurrea, R., 2006. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Information & management, 43(1), pp.1-14.

Forlizzi, J. and Battarbee, K., 2004, August. Understanding experience in interactive systems. In Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (pp. 261-268). ACM.

Foster, M.E., 2007, July. Enhancing human-computer interaction with embodied conversational agents. In International Conference on Universal Access in Human-Computer Interaction (pp. 828-837). Springer, Berlin, Heidelberg.

Finstad, K., 2010. The usability metric for user experience. *Interacting with Computers*, *22*(5), pp.323-327.

Ganglbauer, E., Schrammel, J., Geven, A., & Tscheligi, M. (2009). Possibilities of psychophysiological methods for measuring emotional aspects in mobile contexts. In MobileHCI (Vol. 9, pp. 15-18).

Gaver, W. W. & Martin, H. (2000). Alternatives: Exploring information appliances through conceptual design proposals. Proceedings of the CHI 2000 Conference on Human Factors in Computing (pp. 209-216). New York: ACM Press.

Glanznig, M., 2012. User experience research: Modelling and describing the subjective. Interdisciplinary Description of Complex Systems: INDECS, 10(3), pp.235-247.

Hall, R.J., Snell, A.F. and Foust, M.S., 1999. Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. Organizational Research Methods, 2(3), pp.233-256.

Hamdi, H., Richard, P., Suteau, A., & Allain, P. (2012, June). Emotion assessment for affective computing based on physiological responses. In 2012 IEEE International Conference on Fuzzy Systems (pp. 1-8). IEEE.

Hancock, P.A., Pepe, A.A. and Murphy, L.L., 2005. Hedonomics: The power of positive and pleasurable ergonomics. Ergonomics in design, 13(1), pp.8-14.

Hartmann, J., Sutcliffe, A. and De Angeli, A., 2007, April. Investigating attractiveness in web user interfaces. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 387-396). ACM.

Hassenzahl, M., Platz, A., Burmester, M. and Lehner, K., 2000, April. 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). ACM.

Hassenzahl, M. (2001). The effect of perceived hedonic quality on product appealingness. International Journal of Human-Computer Interaction, 13 (4), 481-499.

Hassenzahl, M., Kekez, R. and Burmester, M., 2002, May. The importance of a software's pragmatic quality depends on usage modes. In Proceedings of the 6th international conference on Work With Display Units (WWDU 2002) (pp. 275-276). ERGONOMIC Institut für Arbeits-und Sozialforschung Berlin.

Hassenzahl, M., 2003. The thing and I: understanding the relationship between user and product. In Funology (pp. 31-42). Springer, Dordrecht.

Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. HumanComputer Interaction, 19, 319-349.

Hassenzahl, M. & Tractinsky, N. (2006). User Experience – a Research Agenda. Behaviour and Information Technology, 25 (2), 91-97.

Hassenzahl, M. and Ullrich, D., 2007. To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. Interacting with computers, 19(4), pp.429-437.

Hassenzahl, M., 2008, September. User experience (UX): towards an experiential perspective on product quality. In IHM (Vol. 8, pp. 11-15).

Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, interactive products – Facets of user experience. Interacting With Computers, 22 (5), 353-362.

Helander, M.G., 2002, September. Hedonomics-affective human factors design. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 46, No. 12, pp. 978-982). Sage CA: Los Angeles, CA: SAGE Publications.

Helander, M.G. and Khalid, H.M., 2006. Affective and pleasurable design. Handbook of human factors and ergonomics, 3, pp.543-572.

Hill, J., Ford, W.R. & Farreras, I.G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. Computers in Human Behavior, 49 (2015) 245–250.

Hsu, S. and Weng, M.H., 2007. An activity-oriented approach to designing a user interface for digital television. In Interactive Digital Television: Technologies and Applications (pp. 148-168). IGI Global.

Hudlicka, E. (2011). Affective Computing: Theory, methods, and applications. CRC Press, Boca Raton.

Ingrosso, A., Volpi, V., Opromolla, A., Sciarretta, E. and Medaglia, C.M., 2015, August. UX and Usability on Smart TV: A Case Study on a T-commerce Application. In International Conference on HCI in Business (pp. 312-323). Springer, Cham.

Isleifsdottir, J., & Larusdottir, M. (2008, June). Measuring the user experience of a task oriented software. In Proceedings of the international workshop on meaningful measures: valid useful user experience measurement (Vol. 8, pp. 97-101).

ISO 9241-210, 2010. Ergonomics of Human–System Interaction – Part 210: Humancentred Design for Interactive Systems.

Jensen, J.F., 2005, November. Interactive television: new genres, new format, new content. In Proceedings of the second Australasian conference on Interactive entertainment (pp. 89-96). Creativity & Cognition Studios Press.

Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K. (2011, March). Physiological signals based human emotion recognition: a review. In 2011 IEEE 7th International Colloquium on Signal Processing and its Applications (pp. 410-415). IEEE.

Jindo, T. and Hirasago, K., 1997. Application studies to car interior of Kansei engineering. International journal of industrial ergonomics, 19(2), pp.105-114.

Jordan, P. W. (2000). Designing pleasurable products. An introduction to the new human factors. London, New York: Taylor & Francis.

Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N. and Stone, A.A., 2004. A survey method for characterizing daily life experience: The day reconstruction method. Science, 306(5702), pp.1776-1780.

Karapanos, E., Zimmerman, J., Forlizzi, J. and Martens, J.B., 2009, April. User experience over time: an initial framework. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 729-738). ACM.

Karapanos, E., 2013. User experience over time. In Modeling users' experiences with interactive systems (pp. 57-83). Springer, Berlin, Heidelberg.

Kari, L. and Rozenberg, G., 2008. The many facets of natural computing. Communications of the ACM, 51(10), pp.72-83.

Kim, Y.H., Kim, D.J. and Wachter, K., 2013. A study of mobile user engagement (MoEN): Engagement motivations, perceived value, satisfaction, and continued engagement intention. Decision Support Systems, 56, pp.361-370.

Kincl, T., & Štrach, P. (2012). Measuring website quality: asymmetric effect of user satisfaction. Behaviour & Information Technology, 31(7), 647-657.

Kempf, D.S., 1999. Attitude formation from product trial: Distinct roles of cognition and affect for hedonic and functional products. Psychology & Marketing, 16(1), pp.35-50.

Kerkow, D., 2007. Don’t have to know what it is like to be a bat to build a radar reflector–Functionalism in UX. Towards a UX Manifesto, p.19.

Kline, R.B., 2011. Convergence of structural equation modeling and multilevel modeling. na.

Kreibig, S.D. (2010). Autonomic nervous system activity in emotion: A review. Biological Psychology, 84, 394–421.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. Interacting with computers, 23(5), 473-483.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K. and Sinnelä, A., 2011, June. Identifying hedonic factors in long-term user experience. In Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces (p. 17). ACM.

Kujala, S., Mugge, R. and Miron-Shatz, T., 2017. The role of expectations in service evaluation: A longitudinal study of a proximity mobile payment service. International Journal of Human-Computer Studies, 98, pp.51-61.

Kurosu, M. and Kashimura, K., 1995, May. Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In Conference companion on Human factors in computing systems (pp. 292-293). ACM.

Kyun Choi, Y., Kim, J. and McMillan, S.J., 2009. Motivators for the intention to use mobile TV: A comparison of South Korean males and females. International Journal of Advertising, 28(1), pp.147-167.

Lachner, F., Naegelein, P., Kowalski, R., Spann, M. and Butz, A., 2016, October. Quantified ux: Towards a common organizational understanding of user experience. In Proceedings of the 9th Nordic Conference on Human-Computer Interaction (p. 56). ACM.

Landis, R.S., Beal, D.J. and Tesluk, P.E., 2000. A comparison of approaches to forming composite measures in structural equation models. Organizational Research Methods, 3(2), pp.186-207.

Law, E.L.C., Roto, V., Hassenzahl, M., Vermeeren, A.P. and Kort, J., 2009, April. 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.

Law, E.L.C. and Van Schaik, P., 2010. Modelling user experience–An agenda for research and practice. Interacting with computers, 22(5), pp.313-322.

Law, E. L. C., & Sun, X. (2012). Evaluating user experience of adaptive digital educational games with Activity Theory. International Journal of Human-Computer Studies, 70(7), 478-497.

LeDoux, J.E., 1995. In search of an emotional system in the brain: leaping from fear to emotion and consciousness.

Lee, S. and Choi, J., 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. International Journal of Human-Computer Studies, 103, pp.95-105.

Lee, B. and Lee, R.S., 1995. How and why people watch TV: Implications for the future of interactive television. Journal of advertising research, 35(6), pp.9-19.

Lekakos, G., Chorianopoulos, K. and Spinelis, D., 2001. Information systems in the living room: A case study of personalized interactive TV design. ECIS 2001 Proceedings, p.59.

Lenderman, M. (2006). Experience the Message: How Experiential Marketing Is Changing the Brand World. New York: Carroll & Graf.

Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A. and Bhogal, R.S., 1997, March. The persona effect: affective impact of animated pedagogical agents. In CHI (Vol. 97, pp. 359-366).

Levenson, R.W. (1992). Autonomic Nervous System Differences Among Emotions. American Psychological Society, 3, pp. 23 – 27.

Lin, J.C.C. and Lu, H., 2000. Towards an understanding of the behavioural intention to use a web site. International journal of information management, 20(3), pp.197-208.

Lindgaard, G. and Whitfield, T.A., 2004. Integrating aesthetics within an evolutionary and psychological framework. Theoretical Issues in Ergonomics Science, 5(1), pp.73-90.

Little, T.D., Cunningham, W.A., Shahar, G. and Widaman, K.F., 2002. To parcel or not to parcel: Exploring the question, weighing the merits. Structural equation modeling, 9(2), pp.151-173.

Little, T.D., Rhemtulla, M., Gibson, K. and Schoemann, A.M., 2013. Why the items versus parcels controversy needn't be one. Psychological methods, 18(3), p.285.

Livaditi, J., Vassilopoulou, K., Lougos, C. and Chorianopoulos, K., 2003, January. Needs and gratifications for interactive TV implications for designers. In 36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the (pp. 9-pp). IEEE.

Luger, E. and Sellen, A., 2016, May. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 5286-5297). ACM.

Luojus, S., 2012, November. Integrating momentary and long-term UX: a theoretical approach. In Proceedings of the 24th Australian Computer-Human Interaction Conference (pp. 353-356). ACM.qui

Macleod, M., 1994. Usability in context: Improving quality of use. In Human Factors in Organizational Design and Management–IV (Proceedings of the International Ergonomics Association 4th International Symposium on Human Factors in Organizational Design and Management, Stockholm.

Mahlke, S. & Thüring, M. (2007). Studying antecedents of emotional experiences in interactive contexts. In Proceedings of the SIGCHI conference on human factors in computing systems. (pp. 915–918). ACM.

Mahlke, S. (2008). User Experience of Interaction with Technical Systems (Doctoral dissertation, Technische Universität, Berlin). Retrieved from https://depositonce.tu-berlin.de/handle/11303/2090

Maia, C. L. B., & Furtado, E. S. (2016, July). A systematic review about user experience evaluation. In International Conference of Design, User Experience, and Usability (pp. 445-455). Springer, Cham.

Malone, T.W., 1981. Toward a theory of intrinsically motivating instruction. Cognitive science, 5(4), pp.333-369.

Mandryk, R. L., Inkpen, K. M., & Calvert, T. W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. Behaviour & information technology, 25(2), 141-158.

Mano, H. and Oliver, R.L., 1993. Assessing the dimensionality and structure of the consumption experience: evaluation, feeling, and satisfaction. Journal of Consumer research, 20(3), pp.451-466.

Maslow, A., 1968. Some educational implications of the humanistic psychologies. Harvard Educational Review, 38(4), pp.685-696.

Matsunaga, M., 2008. Item parceling in structural equation modeling: A primer. Communication Methods and Measures, 2(4), pp.260-293.

McCarthy, J. and Wright, P., 2007. Technology as experience. MIT press.

Mensio, M., Rizzo, G. and Morisio, M., 2018, April. The rise of emotion-aware conversational agents: Threats in digital emotions. In Companion of the The Web Conference 2018 on The Web Conference 2018 (pp. 1541-1544). International World Wide Web Conferences Steering Committee.

Mitchell, T.R., Thompson, L., Peterson, E. and Cronk, R., 1997. Temporal adjustments in the evaluation of events: The “rosy view”. Journal of experimental social psychology, 33(4), pp.421-448.

Morris, S. and Smith-Chaigneau, A., 2012. Interactive TV standards: a guide to MHP, OCAP, and JavaTV. Routledge.

Murnane, K. (2018, May 3). Dumb And Dumber: Comparing Alexa, Siri, Cortana And The Google Assistant. Forbes. Retrieved from https://www.forbes.com/sites/kevinmurnane/2018/05/03/dumb-and-dumber-comparing-alexa-siri-cortana-and-the-google-assistant/#14f1c96436e7

Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with computers*, *22*(5), 336-343.

Nagamachi, M., 1995. Kansei engineering: a new ergonomic consumer-oriented technology for product development. International Journal of industrial ergonomics, 15(1), pp.3-11.

Nagamachi, M., 2002. Kansei engineering as a powerful consumer-oriented technology for product development. Applied ergonomics, 33(3), pp.289-294.

Nass, C, Steuer, J., Tauber, E. & Reeder, H. (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. In INTERACT'93 and CHI'93 Conference companion on Human factors in computing systems, 111-112.

Nakasone, A., Prendinger, H. & Ishizuka, M. (2005). Emotion Recognition from Electromyography and Skin Conductance. Proceedings 5th International Workshop on Biosignal Interpretation (BSI-05).

Nicólas, J.O. and Aurisicchio, M., 2011. The scenario of user experience. In SJ Culley, BJ Hicks, TC McAloone, TJ Howard, and P. Badke-Schaub, Proceedings of The 18th International Conference on Engineering Design (ICED 11) (pp. 182-193).

Norman, D. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.

Novak, T.P., Hoffman, D.L. and Duhachek, A., 2003. The influence of goal-directed and experiential activities on online flow experiences. Journal of consumer psychology, 13(1-2), pp.3-16.

Ogara, S.O., Koh, C.E. and Prybutok, V.R., 2014. Investigating factors affecting social presence and user satisfaction with mobile instant messaging. Computers in Human Behavior, 36, pp.453-459.

Oliver, R.L., 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of marketing research, 17(4), pp.460-469.

Oliver, R.L., Rust, R.T. and Varki, S., 1997. Customer delight: foundations, findings, and managerial insight. Journal of retailing, 73(3), pp.311-336.

Partala, T., Surakka, V. and Lahti, J., 2004, October. Affective effects of agent proximity in conversational systems. In Proceedings of the third Nordic conference on Human-computer interaction (pp. 353-356). ACM.

Picard, R.W. and Daily, S.B., 2005, April. Evaluating affective interactions: Alternatives to asking what users feel. In CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches (pp. 2119-2122). New York, NY: ACM.

Picard, R.W. and Klein, J., 2002. Computers that recognise and respond to user emotion: theoretical and practical implications. Interacting with computers, 14(2), pp.141-169.

Purington, A., Taft, J., Sannon, S., & Taylor, S. (2017). "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Late-Breaking Work, CHI 2017*, May 6–11, 2017, Denver, CO, USA.

Ratner, C. (2000). A Cultural-psychological Analysis of Emotions, Culture and Psychology, 6: 5–39.

Reeves, S. (2017). Some Conversational Challenges of Talking with Machines. *The Economist*, "Now we're talking", Jan 7th 2017. Retrieved from http://www.economist.com/news/leaders/21713836-casting-magic-spell-it-lets-people-control-worldthrough-words-alone-how-voice

Reichheld, F.F., 2003. The one number you need to grow. Harvard business review, 81(12), pp.46-55.

Reiter, U., Brunnström, K., De Moor, K., Larabi, M.C., Pereira, M., Pinheiro, A., You, J. and Zgank, A., 2014. Factors influencing quality of experience. In Quality of experience (pp. 55-72). Springer, Cham.

Rogers, W.M. and Schmitt, N., 2004. Parameter recovery and model fit using multidimensional composites: A comparison of four empirical parceling algorithms. Multivariate Behavioral Research, 39(3), pp.379-412.ca

Roto, V., 2007. User experience from product creation perspective. Towards a UX Manifesto, 31.

Roto, V., Obrist, M., & Väänänen-Vainio-Mattila, K. (2009, August). User experience evaluation methods in academic and industrial contexts. In Proceedings of the Workshop UXEM (Vol. 9, pp. 1-5).

Roto, V., Law, E., Vermeeren, A. and Hoonhout, J., 2011. User Experience White Paper–Bringing clarity to the concept of user experience. Outcome of the Dagstuhl Seminar on Demarcating User Experience, Germany.

Ryu, K., Han, H. and Kim, T.H., 2008. The relationships among overall quick-casual restaurant image, perceived value, customer satisfaction, and behavioral intentions. International Journal of Hospitality Management, 27(3), pp.459-469.

Sauerwein, E., Bailom, F., Matzler, K. and Hinterhuber, H.H., 1996, February. The Kano model: How to delight your customers. In International Working Seminar on Production Economics (Vol. 1, No. 4, pp. 313-327). Innsbruck.

Schrepp, M., Held, T., & Laugwitz, B. (2006). The influence of hedonic quality on the attractiveness of user interfaces of business management software. Interacting with Computers, 18(5), 1055-1069.

Schulze, K. and Krömker, H., 2010, August. A framework to measure user experience of interactive online products. In Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research (p. 14). ACM.

Schütte, S.T., Eklund, J., Axelsson, J.R. and Nagamachi, M., 2004. Concepts, methods and tools in Kansei engineering. Theoretical Issues in Ergonomics Science, 5(3), pp.214-231.

Shin, D.H., 2009. An empirical investigation of a modified technology acceptance model of IPTV. Behaviour & Information Technology, 28(4), pp.361-372.

Sonderegger, A. and Sauer, J., 2010. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. Applied ergonomics, 41(3), pp.403-410.

Spreng, R.A. and Page Jr, T.J., 2003. A test of alternative measures of disconfirmation. Decision Sciences, 34(1), pp.31-62.

Stephenson, M.T. and Holbert, R.L., 2003. A Monte Carlo simulation of observable versus latent variable structural equation modeling techniques. Communication Research, 30(3), pp.332-354.

Thong, J.Y., Hong, S.J. and Tam, K.Y., 2006. The effects of post-adoption beliefs on the expectation-confirmation model for information technology continuance. International Journal of Human-Computer Studies, 64(9), pp.799-810.

Thüring, M. and Mahlke, S., 2007. Usability, aesthetics and emotions in human–technology interaction. International journal of psychology, 42(4), pp.253-264.

Tractinsky, N., Shoval-Katz, A. & Ikar, D. (2000). What is beautiful is usable. Interacting with Computers. 13(2), 127-145.

Tractinsky, N., Shoval-Katz, A. & Ikar, D. (2000). What is beautiful is usable. Interacting with Computers. 13(2), 127-145.

Tuch, A.N., Roth, S.P., HornbæK, K., Opwis, K. and Bargas-Avila, J.A., 2012. Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. Computers in Human Behavior, 28(5), pp.1596-1607.

Van Schaik, P., & Ling, J. (2008). Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. Interacting With Computers, 20(3), 419–432.

Vermeeren, A. P., Law, E. L. C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010, October). User experience evaluation methods: current state and development needs. In *Proceedings of the*

Venkatesh, V., 2000. Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. Information systems research, 11(4), pp.342-365.

Walsh, T., Varsaluoma, J., Kujala, S., Nurkka, P., Petrie, H. and Power, C., 2014, November. Axe UX: Exploring long-term user experience with iScale and AttrakDiff. In Proceedings of the 18th International Academic MindTrek

Conference: Media Business, Management, Content & Services (pp. 32-39). ACM.

Wang, Y. J., Hong, S., & Lou, H. (2010). Beautiful beyond useful? The role of web aesthetics. Journal of Computer Information Systems, 50(3), 121-129.

Weniger, S., 2010, June. User Adoption of IPTV: A Research Model. In Bled eConference (p. 30).

Wilson, G.M. & Sasse, M.A. (2000b). Investigating the Impact of Audio Degradations on Users: Subjective vs. Objective Assessment Methods. In OZCHI 2000: Interfacing Reality in the New Millennium, pp. 135 – 142(Sydney, Australia).

Wirtz, D., Kruger, J., Scollon, C.N. and Diener, E., 2003. What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. Psychological Science, 14(5), pp.520-524.

Wright, P. and Blythe, M., 2007, September. User experience research as an inter-discipline: Towards a UX Manifesto. In E. Law, A. Vermeeren, M. Hassenzahl, & M. Blythe (Eds.), Proceedings of the Workshop on Towards a UX Manifesto (pp. 65-70).

Yang, C., Nay, S. and Hoyle, R.H., 2010. Three approaches to using lengthy ordinal scales in structural equation models: Parceling, latent scoring, and shortening scales. Applied Psychological Measurement, 34(2), pp.122-142.

Yao, L., Liu, Y., Li, W., Zhou, L., Ge, Y., Chai, J., & Sun, X. (2014, June). Using physiological measures to evaluate user experience of mobile applications. In International Conference on Engineering Psychology and Cognitive Ergonomics (pp. 301-310). Springer, Cham.

Youngblood, N.E. and Youngblood, S.A., 2013. User experience and accessibility: An analysis of county web portals. Journal of Usability Studies, 9(1), pp.25-41.

Yuan, M. (2016, August 15). The conversational chatbox design challenge: Learn why chatbots can fail and what can be done to make them successful.

Retrieved from https://developer.ibm.com/articles/cc-cognitive-chatbot-conversational-design/

Zhang, P. and Li, N., 2004. Love at first sight or sustained effect? The role of perceived affective quality on users' cognitive reactions to information technology. ICIS 2004 proceedings, p.22.

Zhang, P. and Von Dran, G.M., 2000. Satisfiers and dissatisfiers: A two-factor model for website design and evaluation. Journal of the American society for information science, 51(14), pp.1253-1268.

Appendix 1.

< Before UX questions >

Usability (Finstad, 2010).

- U1. Using the service is (will be) a frustrating experience.
- U2. The service is (will be) easy to use.
- U3. I will need to spend too much time correcting things with this service.
- U4. The service will meet my requirements.

Enjoyment (Mitchell et al., 1997; Wirtz et al., 2003).

- E1 I will enjoy using the service.
- E2 I think using the service will be fun.
- E3 I think using the service will be rewarding.
- E4 I will be satisfied with the service.

< After UX questions >

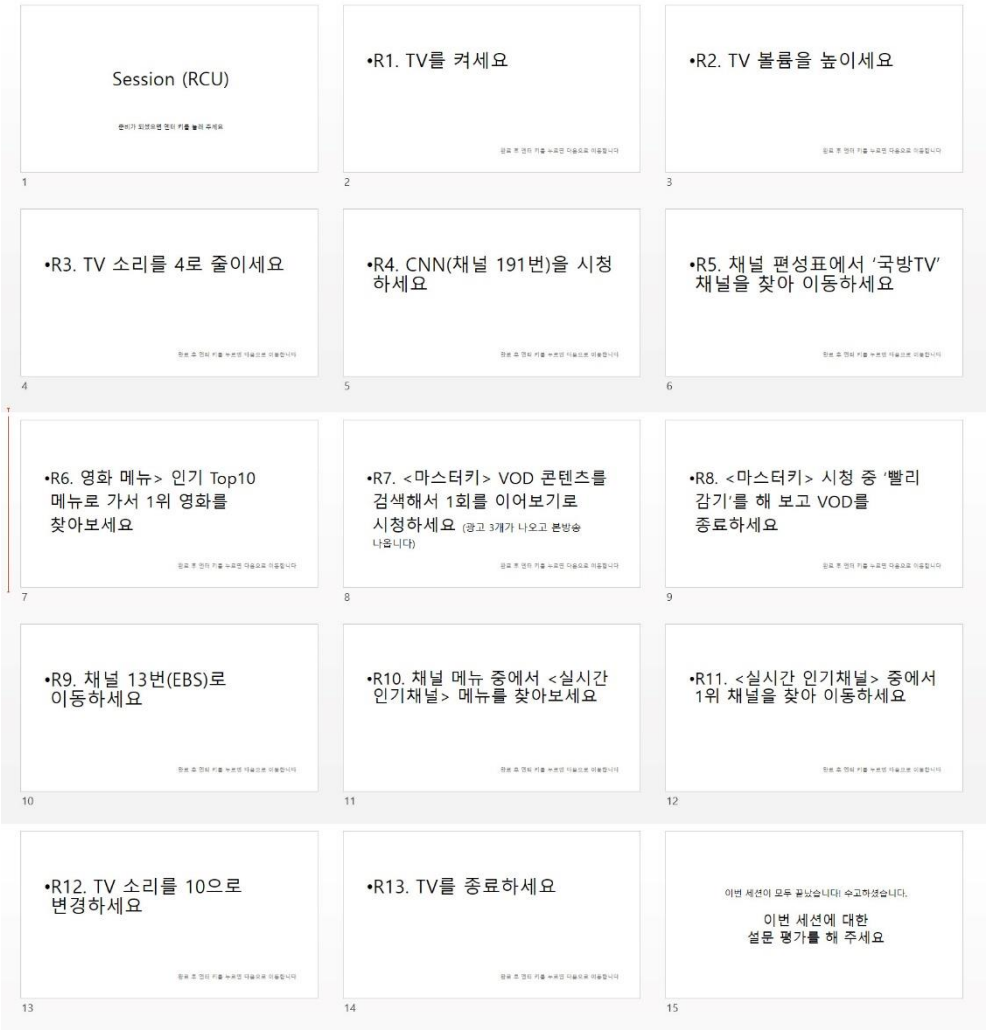
Behavioral intentions (Reichheld, 2003; Wirtz et al., 2003).

- B1. Would you start using the service again (assuming you hadn't just used it, but that you know what you now know)?

- B2. Based on your experience, how willing are you to continue using the service?
- B3. How likely is it that you would recommend the service to a friend who is interested in it?
- B4. How likely is it that you will be using the service in the future?

Appendix 2.

Physiological measurements task sheets



Session (AI)

준비가 되셨으면 연속 키를 눌러 주세요

16

•A1. TV를 켜세요

완료 후 연속 키를 누르면 다음으로 이동합니다

17

•A2. TV 볼륨을 높이세요

완료 후 연속 키를 누르면 다음으로 이동합니다

18

•A3. TV 소리를 6으로 줄이세요

완료 후 연속 키를 누르면 다음으로 이동합니다

19

•A4. CNBC(채널 197번)을 시청하세요

완료 후 연속 키를 누르면 다음으로 이동합니다

20

•A5. '바둑TV' 채널을 찾아 이동하세요

완료 후 연속 키를 누르면 다음으로 이동합니다

21

•A6. '강동원' 나오는 영화를 찾아보세요

완료 후 연속 키를 누르면 다음으로 이동합니다

22

•A7. <거기가 어딘데> VOD 콘텐츠를 검색해서 1회를 이어보기로 시청하세요 (광고 3개가 나오고 본방송 나옵니다)

완료 후 연속 키를 누르면 다음으로 이동합니다

23

•A8. <거기가 어딘데> 시청 중 '앞으로 점프'를 해 보고 VOD를 종료하세요

완료 후 연속 키를 누르면 다음으로 이동합니다

24

•A9. 채널 11번(MBC)로 이동하세요

완료 후 연속 키를 누르면 다음으로 이동합니다

25

•A10. 로맨스 영화를 검색해 보세요

완료 후 연속 키를 누르면 다음으로 이동합니다

26

•A11. 로맨스 영화 결과 중에서 첫번째 영화를 찾아 상세정보 화면으로 이동하세요

완료 후 연속 키를 누르면 다음으로 이동합니다

27

•A12. TV 소리를 10으로 변경하세요

완료 후 연속 키를 누르면 다음으로 이동합니다

28

•A13. TV를 종료하세요

완료 후 연속 키를 누르면 다음으로 이동합니다

29

이번 세션이 모두 끝났습니다! 수고하셨습니다.

이번 세션에 대한
설문 평가를 해 주세요

30

Appendix 3.

< Attrakdiff[2]> questionnaires for IPTV UX (Hassenzahl et al., 2010)

Date _____ Sub. No. _____

아래 각 문장을 잘 읽고 현재 자신의 기분상태와 가깝다고 생각되는 정도를 -3 ~ 3 중에서 선택해 주십시오

	-3	-2	-1	0	1	2	3	
기계적인								인간적인
복잡한								단순한
비실용적인								실용적인
성가신								간단한
예측하기 어려운								예측 가능한
혼란스러운								명확한
휘어잡을 수 없는								통제 가능한
단절된								연결된
비전문적인								전문적인
초라한								세련된
저렴한								값비싼
분열시키는								통합하는
나를 단절시키는								나를 가깝게 연결 시키는
표현할 수 없는								표현할 수 있는
기존의								새로운
상상력이 없는								창의적인
조심스러운								대담한
보수적인								혁신적인
지루한								매력적인
요구가 많지 않은								도전적인
평범한								귀한
즐겁지 않은								즐거운
추한								매력적인
불쾌한								유쾌한
거부하는								호응하는
나쁜								좋은
반발하는								호소하는
실망시키는								동기를 부여하는

Appendix 4.

DRM questionnaires sheet

Sub. No. _____

TV 사용 경험에 관한 사용자 조사

안녕하세요. 서울대학교 산업공학과 휴먼인터페이스 연구실입니다. 본 설문에 응해 주셔서 감사합니다. 저희는 우리 일상에서 가장 친숙한 기기 중 하나인 TV의 사용 경험에 대해 총 4주에 걸쳐 알아보고자 합니다.

바쁘시더라도 잠시만 시간을 내시어 매일 일정한 시간에 TV 사용 내용을 기록해 주세요. 작성 방법의 예시는 아래 내용을 참고해 주시기 바랍니다. 작성 방법을 숙지하신 뒤, 다음 페이지의 기초 문항을 거쳐 3페이지부터 준비된 양식에 따라 4주간 매일 1페이지씩 작성해 주시면 됩니다. 처음에는 다소 시간이 걸릴 수 있으나 익숙해지면 10~15분 이내로 작성하실 수 있습니다.

부탁드릴 말씀은 가급적 1일 1회 작성을 원칙(시간이 지나면 잘 기억나지 않기 때문에)으로 지켜 주시는 것과, 솔직하고 구체적으로 작성해 주시는 것입니다. 본 연구는 실생활에서 나타나는 생생한 TV 경험 데이터를 통해 TV 사용 실태를 파악하고 이를 통해 가치 있는 미래 TV 서비스를 만드는데 기초 자료로 사용될 예정입니다. 또한 작성해주신 자료는 본 연구 외에 어디에도 공유되지 않을 것임을 약속 드립니다.

.

<Workbook 작성방법>

(1) 매일 1회, 그 날 저녁이나 다음날 아침 작성합니다 (가급적 같은 시간)

(2) 시청 시간을 기록합니다 (여러 번 나눠서 시청한 경우 가장 기억에 남거나 마지막 시청한 내용을 중심으로 작성합니다)

(3) 시청한 본인 및 같이 있었던 구성원을 모두 선택합니다

(4) TV를 시청하면서 사용한 기기를 선택합니다

(5) 시청 동기와 내용을 자유롭게 작성합니다

(예 1) "청소하면서 자연스럽게 TV를 켜 놓음. 굳이 TV 를 보는 건 아니지만 청소하면서 아침 드라마를 본다 "

(예 2) "기가지니한테 TV를 켜 달라고 함. 리모컨을 찾지 못해도 음성으로 켤 수 있어 편리했다"

(예 3) " 아는 형님을 보다가 내일 날씨가 궁금해서 지니한테 물어봤으나 말을 알아듣지 못해 짜증이 남.."

(6) TV를 시청할 때 기분 (feeling)을 세분화해서 평가합니다. 자신의 기분과 매우 비슷하면 6점, 조금 비슷하면 4나 5점, 반반이면 3점, 거의 다르면 1이나 2점, 전혀 다르면 0점을 주시면 됩니다

(예) '행복하다 ' 항목에서 내 상태가 '어느 정도' 행복하다 면 4점이나 5점

'지친다' 항목에서 내 상태가 '전혀 지치지 않았다' 면 0점

	전혀 아니다			보통			매우 그렇다
행복하다	0	1	2	3	4	5	6
지친다	0	1	2	3	4	5	6

Part 1-1. 본 설문에 들어가기 전에 아래의 사전 질문에 대해 응답해 주시기 바랍니다.
(Part 1-1과 1-2는 한 번만 작성해 주시면 됩니다)

- **귀하의 성별을 선택해 주세요** ① 여자 ② 남자

- **귀하의 연령은 어떻게 되시나요?** _____ 세

- **현재 귀하의 가족 구성원은 모두 몇 명 입니까?**

① 1명 ② 2명 ③ 3명 ④ 4명 ⑤ 5명 ⑥ 6명 이상

- **평소 하루에 TV를 시청하는 시간은 몇 시간 정도입니까?**

① 1시간 내외 ② 1 ~ 2시간 내외 ③ 2 ~ 3시간 내외
④ 3 ~ 4시가 내외 ⑤ 4시간 이상 ⑥ 하루 종일
⑦ 거의 시청하지 않음

- **현재 귀하의 댁에 설치되어 있는 TV는 어느 통신사에서 제공하는 상품입니까?**

① KT ② LG U+ ③ SK 브로드밴드 ④ 기타 _____

- **현재 귀하의 댁에 설치되어 있는 TV는 음성 인식 기능을 제공하는 상품입니까?**
(음성인식 셋탑박스 또는 음성인식 기능이 지원되는 리모컨 포함)

① 예, 음성인식 기능을 제공하는 스마트 셋탑박스입니다
② 예, 리모컨에 음성인식 기능이 있습니다
③ 예, 셋탑박스와 리모컨 모두 음성인식 기능을 제공합니다
④ 아니오, 둘 다 기능이 없습니다
⑤ 잘 모르겠습니다

체크하신 후에 다음 페이지로 이동해주시기 바랍니다.

Part 1-2. 본 설문에 들어가기 전에 아래의 사전 질문에 대해 응답해 주시기 바랍니다.
(Part 1-1과 1-2는 한 번만 작성해 주시면 됩니다)

- **이 서비스를 사용하면서 실망할 것 같다**

전혀 동의하지 않는다 매우 동의한다
0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6

- **이 서비스를 사용하기 쉬울 것 같다**

전혀 동의하지 않는다 매우 동의한다
0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6

- **이 서비스를 사용하는데 너무 많은 시간을 쏟을 것 같다**

전혀 동의하지 않는다 매우 동의한다
0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6

- **이 서비스는 내 요구사항을 충적시켜 줄 것 같다**

전혀 동의하지 않는다 매우 동의한다
0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6

- **나는 이 서비스를 즐길 것 같다**

전혀 동의하지 않는다 매우 동의한다
0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6

- **나는 이 서비스가 재미있을 것 같다**

전혀 동의하지 않는다 매우 동의한다
0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6

- **나는 이 서비스 사용이 보람있는 일이라고 느낄 것 같다**

전혀 동의하지 않는다 매우 동의한다
0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6

- **나는 이 서비스에 만족할 것 같다**

전혀 동의하지 않는다 매우 동의한다
0 ---------- 1 ---------- 2 ---------- 3 ---------- 4 ---------- 5 ---------- 6

Part 1이 모두 완료되었습니다. 이제 다음 페이지로 이동해서 오늘의 TV 사용 경험에 대해 기록해 주시기 바랍니다

- **TV 시청시간** 2018년 월 일 시 분부터 ~ 시 분까지

- **TV 시청할 때 누구와 함께 있었나요? (여러 명이 함께 시청한 경우 모두 선택)**

☐ 혼자 ☐ 배우자 ☐ 자녀 ☐ 부모님 ☐ 형제 / 친척 ☐ 이웃 등 기타

- **TV 시청하면서 함께 사용했던 기능(또는 기기)를 모두 선택해 주세요 (여러 개일 경우 모두 선택)**

☐ 리모컨 ☐ 음성인식 지원 셋탑 또는 리모컨의 음성인식기능

☐ 동작인식기능 ☐ 스마트폰 ☐ 태블릿

☐ PC 또는 노트북 ☐ 기타: _____

- **TV 시청 경험 (함께 있던 사람, 사용한 기능/기기, 시청 내용, 감정 등)을 자유롭게 적어주세요. 가능한** 음성인식 기능을 활용한 경험 위주로 적어 주시면 좋습니다.

- **TV를 시청하는 동안 기분(feeling)은 어떠셨나요? 아래 각 항목 (① 번부터 ⑩ 번까지) 마다 나의 기분과 가장 가까운 쪽에 체크해 주세요. (전혀 못 느꼈다 0 ..... 매우 강하게 느꼈다 6)**

	전혀 못 느꼈다						매우 강하게 느꼈다
① 유능한 / 자신 있는	0	1	2	3	4	5	6
② 지루한	0	1	2	3	4	5	6
③ 행복한	0	1	2	3	4	5	6
④ 긴장한 / 스트레스 받는	0	1	2	3	4	5	6
⑤ 흥미로운 / 열중하는	0	1	2	3	4	5	6
⑥ 우울한 / 울적한	0	1	2	3	4	5	6
⑦ 편안한	0	1	2	3	4	5	6
⑧ 짜증나는 / 화가 나는	0	1	2	3	4	5	6
⑨ 우호적인 / 애정이 있는	0	1	2	3	4	5	6
⑩ 피곤한	0	1	2	3	4	5	6

- **TV 시청 후 나의 기분은 어땠나요? 솔직하게 자기 평가해 주세요.**

① 좋다 ② 보통이다/잘 모르겠다 ③ 나쁘다

오늘 작성은 모두 마무리되었습니다. 감사합니다

(일주일마다) 오늘은 지난 일주일간 TV 사용 관해 몇 가지만 더 응답해 주시기 바랍니다.

- **현재 사용하시는 TV 서비스(KT 올레tv, LG U+ 등)가 필요하다고 생각하십니까?**

	전혀 동의하지 않는다					매우 동의한다
필요하다	0 ----------	1 ----------	2 ----------	3 ----------	4 ----------	5 ---------- 6

- **현재 사용하시는 TV 서비스(KT 올레tv, LG U+ 등)에 만족하십니까?**

	전혀 동의하지 않는다					매우 동의한다
만족한다	0 ----------	1 ----------	2 ----------	3 ----------	4 ----------	5 ---------- 6

- **현재 사용하시는 TV 서비스(KT 올레tv, LG U+ 등)을 앞으로도 계속 사용할 예정이십니까?**

	전혀 동의하지 않는다					매우 동의한다
계속 사용한다	0 ----------	1 ----------	2 ----------	3 ----------	4 ----------	5 ---------- 6

오늘 작성은 모두 마무리되었습니다. 감사합니다

Appendix 5.

User comments in field test (about CA and RCU)

Day	User comments	Emotion
D1	기가지니 셋탑이 갑자기 동작을 안하고 불러도 대답이 없음. 리모컨 없이 tv를 보다 갑자기 리모컨을 찾는게 불편했음. 누워서 말로 tv를 켜고 끄는 게 익숙해지면서 편한거 같다	Positive
	아기가 있기 때문에 일어나면 tv로 애니메이션을 틀어줍니다. 리모컨을 아기가 들고다니면서 막 누르기 때문에 다른 곳에 고정으로 두고 거의 대부분을 음성인식을 이용합니다. 항상 티비를 틀어놓기 때문에 리모컨보다는 음성인식을 종일 사용합니다. 아기가 보고 싶어하는 애니메이션을 말로 바로 찾아주기에 매우 편합니다.	Positive
	오전에 아침을 먹으면서 야구를 보았고 기가지니에게 아기 상어 노래와 팟케스트를 잠시 틀어 달라고 했었다.	Neutral
	저녁을 먹으면서 틀어놓고 있었는데 얘기를 나누다가 지니가 잘 못 인식했는지 채널이 예기치 않게 넘어갔다	Negative
	리모컨으로 TV를 켬. 아는 형님 재방송을 시청함 시청 중 막내가 음성으로 레이디버그 검색. 유트브로 검색창 열림. 각국 언어로 업데이트한 것 중 두 편을 골라서 시청. 아이들에게 리모컨, 혼자 스마트폰 검색이 끝없이 계속 연결되어 시청을 하게 되서 강제로 시청 종료함	Neutral
D2	농구 채널을 틀어 달라고 했지만 지니가 찾지 못해서 결국 리모콘을 사용함. 정확한 번호나 방송사명을 이야기하지 않으면 안되는 것 같음	Negative
	기가지니에서 아기 상어를 찾아 달라고 해서 들으면서 빨래를 접고 딸아이는 티비를 시청했어요.	Neutral
	하루종일 미국 프로야구를 봤는데 리모콘으로 TV 전원만 쓰면 기가지니가 인식하지 못하는걸 알았다.. 기가지니를 통해 홈쇼핑주문 하는걸 공부 해야겠다	Neutral
	집에오자마자 일단 기가지니에게 tv를 틀어달라고 함. 리모컨 찾기 전에 먼저 tv를 켜는게 편리하다고 봄. 보다가 다른 채널로 돌릴 땐 다기 리모컨을 찾는다. 결국 두 가지 다 사용중	Positive
	음성인식으로 티비를 켜고 아기에게 음성인식로 동화를 틀었습니다. 아기가 CF 보는걸 좋아해서 리모콘으로 이곳저곳 돌리며 채널을 찾았습니다. TV에서 나오는 소리들 중 '지니'라는 말이 나오면 AI 스피커가 반응하여 갑자기 다른것이 실행되었습니다	Negative
	늦게 퇴근을 했는데 신랑이 골프채널을 보고 있어서 옆에서 같이 시청했다. 중간에 내가 원하는 채널로 음성명령을 하여 채널 변경을 하	Positive

	는 등 장난을 쳤다.	
	Tv를 리모콘으로 틀어놓고 집안일을 하다가 안마의자에 앉아서 안마를 받고 있었는데 보던 프로그램이 끝났다. 안마의자 사용 중이라 일어나지 못하는 상황이라 지니를 불러서 실시간 인기채널을 검색하고 다른 채널로 이동하였다.	Positive
	리모컨으로 TV 켜 놓고 AI에게 미세먼지 체크. 이 날은 TV 켠 후 계속 스마트폰만 검색하다 잠이 듦	Neutral
D3	티비 음성인식으로 원하는 프로그램 찾기를 시도했으나 인식정확도가 떨어져서 포기하고 그냥 어머니와 보고 있던 드라마를 계속 봤다.	Negative
	신랑과 저녁을 먹고 들어오자마자 쇼미더머니를 보기 위해 tv를 켰다. 본 방송 시작 전까지 시간이 남아 음성으로 기가지니 롯데슈퍼 장보기 서비스를 이용	Neutral
	볼게 없어서 새로나온 영화 검색(음성)하고 그 중에 신과함께 시청함. 음성인식이 어떨땐 잘 되고 어떨땐 몇 번 불러도 못알아듣는 경우가 있어 익숙해지기 힘듦. ㅠㅠ	Neutral
	음성인식으로 '옹성우 보여줘'라고 했지만 이름이 특이해서인지 몇차례의 시도 끝에야 검색 할 수 있었음. Tv와 핸드폰을 동시에 하다가 옹성우가 주인공으로 내정된 드라마의 상대역으로 김향기배우가 낙점되었다는 기사를 읽고 김향기 배우의 영화나 드라마를 보고 싶어 '김향기 보여줘'라고 했지만 계속 실패해 그냥 리모콘으로 검색함	Negative
	한국시리즈 야구 결과가 궁금해서 기가지니에게 야구를 틀어달라고 하니 역시나 못 알아들음. 리모컨으로 내가 켜서 하이라이트를 보는데 다행히 SK가 이겼다.	Negative
	음성지원은 자주 오류가 난다. 계속 듣는중이라고 뜨고 종료되지도 않는다. TV를 자주 껐다 켬.	Negative
D4	어머니께 음성인식 검색 기능을 알려 드림. 숨바꼭질을 검색하는 것까지는 가능하셨지만 세부적인 선택、 이어보기 등을 조작하는 것을 몇번하다가 포기하셨다. 나도 그런 세부조작은 리모콘과 병행하는게 편하다고 느꼈다.	Neutral
D5	오늘 tv를 보면서 기가지니로 채널을 돌리려 했지만 노래가 틀려서 당황해서 얼른 꺼버렸다. 부모님께서는 음성인식을 쓰면 노래가 자주 틀려지는데、 언제는 노래를 틀까요?하고 선택지가 나와서 취소하기 쉽지만 언제는 바로 노래가 나와서 짜증이 난다고 말씀하셨다.	Negative
	음성인식으로 TV 끔	Neutral
	기가지니 음성인식이 정확치 않다고 느껴졌다. 발음의 문제인지, 발성의 문제인지... 음성인식은 TV틀어줘, TV꺼줘, 리모컨 찾아줘 이런 용도로 가장 편한듯하다	Negative
	혼자 TV 시청 시 음성으로 리모컨 기능 찾기는 정말 필요함. 오늘도 TV 시청 못할 뻔 함	Positive
D6	지니를 불러도 안되서 리셋 후에.작동을 다시 시킴. 이유를 모르겠으나 지니가 작동하지않음. 전원 오프 온 후 다시 잘됨	Negative

	혼자서 줄리아로버츠 주연의 영화를 보다가 음성기능 지니한테 수유음악 틀어 달라고 해서 수유를 한 후 음악을 꺼달라고 하고 영화를 이어 봤다. 수유하는 동안 음성인식이 일을 해주니 너무 좋다고 생각했다.	Positive
	TVN 정액제로 미스터선샤인을 시청. 음성 인식으로 티비를 켜고 싶으나 티비와 셋탑이 연동이 안되어 따로 켜야됨을 깨닫고 포기	Negative
	야구경기 결과가 궁금해서 탑재된 음성인식 기능으로 “오늘 야구경기 알려줘”라고 하고 결과를 들었다	Positive
	일어나서 리모컨을 이용하여 티비를 켜고 음성인식으로 아기에게 뽀로로를 틀어줌. 티비를 보던 중 갑자기 음성인식이 되지않아 셋탑을 리셋	Neutral
	빨래를 널면서 핸드폰으로 보던 유투브영상을 보려고 음성인식으로 검색을 하였으나 방송을 검색할 때보다 왠지 더 알아듣지 못했다. 해당 단어 발음의 문제인가 해서 즐겨보던 다른 것도 유투브 검색을 시도했지만 살패했다. 아마 방송검색보다 유투브 검색이 고유명사 같은 것의 폭이 넓어서 그런가 하고 추측할 뿐이었다.	Negative
	영화를 보면서 중간에 지니한테 시간을 물어 봄	Neutral
	음성인식으로 레이디버그검색. 항상 유투브가 먼저 검색된다. 시청 중간 중간 음성인식으로 다른 주인공 이름 검색. 마드리앙 이름으로 계속 다른 인물만 검색됨. 막내의 영어 발음은 검색이 잘 안되는 것 같다	Neutral
D7	날씨가 흐린데 일기예보를 보지는 못해서 ai 기능으로 날씨를 물어보았음.	Neutral
	집안일을 하면서 옹성우가 나오는 프로그램을 보려고 음성인식을 시도했는데 한번에 되어서 의아해하면서도 기분 좋게 보았다. 검색한 이력이 있어서 쉽게 검색이 되는건가 생각해보았다.	Positive
	리모컨을 찾지 못해 기가지니에게 리모컨을 찾아 달라고 함.	Positive
	음성인식으로 밥블레스유 11화 검색. 항상 느끼지만 회차까지 말하면 잘 못알아듣는다. 밥블레스유 검색 후 리모컨으로 원하는 회차로 이동해서 시청함	Negative
D8	지니에게 11번을 틀어달라 했는데 인식이 안되서 엠비씨라고 하니 인식됨	Negative
	밥을 먹으면서 기가지니에게 티비를 켜달라고 이야기를 했다. 리모컨을 못찾아 기가지니에게 찾아 달라고도 했다.	Positive
	음성인식으로 티비를 켜고 지니에게 날씨를 물어봄. 리모컨으로 신서유기를 찾아 시청함. 아기가 뽀로로를 보는 중에 리모컨을 건드려 소리가 커져서 지니에게 줄여달라고 했음.	Positive
	해외 슬라임"영상 검색.. 유투브에 채널이 너무 많아 이것 저것 리모컨으로 검색한다. 자료가 많이 있어 음성으로 검색 후 다시 리모컨으로 조작하는데 음성인식이 연결이 안되어 검색을 못했다.	Negative
	다같이 저녁을 먹으면서 신서유기 시청 리모콘을 못찾아 기가지니에	Positive

D9	게 찾아 달라고 했고 기가지니가 알아서 다음편도 말하면 틀어줘서 좋다고 가족끼리 의견을 나눔	
	기가지니로 동화를 아이에게 읽어주었다. 음성인식으로 드라마를 찾았지만 리모컨 없이는 조작이 어려워 볼 수 없었다.	Negative
	늦게 들어와서 습관적으로 tv를 켜고 피곤한 상태로 뉴스 시청. 꼼짝도 하기 싫어서 말로 채널 좀 돌려달라고 했는데 계속 못알아듣고 엉뚱한 걸 띄워서 짜증이 남	Negative
	오랜만에 음성인식으로 미세먼지 체크."요즘 새로 하는 드라마"로 음성 검색하니 잘 못알아들어서 다시 "최신 드라마 알려줘"라고 검색함	Neutral
	"악마는 프라다를 입는다" 시청 도중 갑자기 생각난 영화제목을 지니에게 음성 검색했으나 해당 영화가 없던 것인지 못 알아들은 건지 검색되진 않았다	Negative
D10	농구를 보면서 60번 61번을 왓다갔다를 지니에게 시켰는데.잘됬습니다	Positive
	유투브로 아이에게 동요를 틀어줬음. 지니에게 TV를 켜달라고 했는데 셋탑만 꺼지고 TV 본체는 꺼지지 않았다.	Negative
	예능 프로그램을 시청하는데 대사를 잘못 인식하여 음성인식이 작동한 경우가 두 번 있어 잠시 불편함을 느낌	Negative
	AI스피커(기가지니)로 티비를 켜고 리모컨으로 실시간 인기채널 검색	Neutral
	친언니와 tv시청을 하다가 기가지니로 시간,날씨 등등 여러가지 기능도 시켜보았다. 신기하고 편리한 점도 있고 답답할 때도 있다. 우리집 기가지니 이름은 '친구'로 저장되어있는데, 친구는 말을 잘 못 알아들을 때가 많아서 답답했다	Neutral
	아이들 나라 리모컨 검색보다 음성인식으로 검색하는 것이 더 쉬운 듯. 아이들이 많이 사용함	Positive
D11	음성인식 스피커는 사용하지 않았어요. 이유는 아무리 불러도 대답은 안했어요.	Negative
	신랑과 설악산1박 여행을 가서 호텔에서 잤다. 티비를 음성으로 사용하다 오랫만에 리모컨으로만 작동하려니 좀 불편했다	Neutral
	애기를 안으면서 AI스피커로 볼륨 조절이 가능해서 편리했다. 키고 끄는 기능은 티비가 꺼지지 않고 셋탑박스만 꺼져서 아쉽다.	Positive
	음성인식으로 30분 뒤 꺼짐 예약을 설정할 수 있다고 듣고 그대로 해 봤는데 잘 예약이 되어서 기분좋았다	Positive
	기가지니의 성능에 실망이다.. 단순한 기능만 가능한 듯 하다.	Negative
	TV 리모컨으로 켬. 음성인식으로 동백 배달 맛집 검색함. 음성인식은 검색이 제일 편하고 검색 후에 결과에 대한 명령은 잘 안된다. 특정 메뉴로 다시 검색한 뒤 스마트폰으로 리뷰 뒤짐.	Neutral
	지니에게 리모컨 찾아달라고 함	Neutral
D12	음성인식으로 영화 제목 (리틀 포레스트)을 검색해서 뜨개질을 하면서 시청함	Neutral
	지니에게 이것 저것 시켜 봤는데 내 음성톤에 따라 조금만 높아지면	Neutral

	말을 못 알아듣는게 확실함	
	기가지니에게 TV 틀어줘,리모컨 찾아줘 연속 실행	Neutral
D13	1박2일을 보면서 음성인식기능을 사용. 볼륨을 많이 내려달라는 말을 인식하지 못함 볼륨 시켜줘 라고 하면서 잘못들었다고 함	Negative
	기가지니에게 음악을 추천해달라고 해서 음악을 들으며 딸아이 우유를 먹인다 먹이고 나서 아침 먹은걸 치우며 기가지니에게 뉴스를 틀어 달라고 했다.	Neutral
	평소에 자주 찾던 실시간 인기채널을 틀었다. 인기채널 중 마음에 드는 채널을 찾더니 지니에게 5위 틀어줘 했지만 지니가 알아듣지 못했다. 전에도 생각한 적 있지만 지니는 맥락을 좀 더 파악 할 수 있었으면 좋겠다 ㅠㅠ	Negative
	TV 켠 후 아이들 보고 싶은 채널을 음성검색으로 찾아 시청. 늘 해외 유투브 채널이나 애니메이션 검색해서 봄. 미성년자 보호 설정이 안되어 있어 옆에 앉아 스마트폰을 하면서 함께 시청함	Neutral
	시청 중 아이들과 얘기중이었는데 갑자기 기가지니가 "써니 사이드업은 어쩌구 저쩌꾸.." 혼자말을 해서 깜짝 놀랐다	Negative
D14	날씨를 놓쳐서 AI스피커로 날씨를 확인	Neutral
	지니에게 동요를 틀러달라고 하여 아기와 함께 들었다.	Neutral
	항상 기가지니로 티비를 켜는데 역시나 티비 온오프 용이지 사용성이 매우 떨어진다. 리모컨이 편하다..	Negative
	시청 중 수면모드. 별다른 움직임이 없으면 기가지니가 "시청을 종료할까요?"물어보고 대답이 없으면 종료했으면 함	Neutral
D15	7개월된 아기가 있는데 기가지니에 도움을 많이 받아요. 수유시에는 두 손을 쓸 수가 없어서 배이비티비 틀어줘 했더니 틀어줘서 무척 기특했습니다.	Positive
	기가지니로 티비를 켜놓고 라디오처럼 티비를 들으며 할 일을 했다. 기가지니가 말로 도와주니 고마운데 대답을 안 할 땐 정말 짜증난다. 잘 안들리나 보다.	Neutral
	지니한테 소리를 줄여달라고 했다. TV를 켜고 끌 땐 점점 더 음성인식으로 하는 게 익숙해지고 있다.	Positive
	음성인식으로 어거스트 러쉬 검색. 영화 시청 중 프레디 하이모어가 나온 영화를 다시 음성 검색함. 음성검색 시 긴 문장은 끊어 읽어줘야 잘 알아듣는 듯 하다.	Neutral
D16	가수를 음성으로 검색하거나 뮤직비디오를 음성검색하면서 시청. "기가지니, 블랙핑크 제니가 나왔던 프로그램 찾아줘"라고 해 보았으나 전혀 관련없는 것들이 잔뜩 나옴. 괜히 했다 싶어 후회했다	Negative
D17	아들과 함께 리모콘을 이용했다. 기가지니를 사용한 건 티비를 꺼달라고 할 때다.	Neutral
	“정식으로” 라고 말했는데 갑자기 지니가 대답함	Negative
	막내가 리모컨으로 tv를 켬. 음성인식으로 '놓지마 정신줄' 검색. 역시 끊어서 읽어야 검색이 한번에 됨.	Positive

D18	갑자기 셋탑을 인식하지 못해서 여러 번 전원을 껐다 켬.	Negative
	기가지니로 일어나서 음악을 틀어달라고 했고 남편이 나와 기가지니에게 뉴스를 틀어달라고 했다.	Neutral
	기가지니에게 말을 걸 때 한 번 꼬이면 절대 못 알아 듣는 경우가 있다.'확인'이라는 말을 여러 번 해도 못 알아 들을 때 짜증이 난다.	Negative
	"기가지니, 맛있는 음식점 찾아줘" 한번 시험삼아 해봤는데 "앱에서 우리집 주소를 등록해주세요"라고 답함 -> 주소를 등록하면 맛집을 보여주나?	Neutral
D19	아침 일찍 등산준비를 하면서 뉴스를 틀어놓음. 일기예보 확인을 위해 ai스피커로 날씨를 확인함	Neutral
	기가지니에게 아기 수유하며 예능프로그램을 틀어달라고 부탁함 말로 다 해결을 해주니 아기에게 수유하며 티비를 볼 수 있어서 너무 좋았다.	Positive
	아는형님' 시청. 시청 중 기가지니한테 "오늘 미세먼지 어때?"하고 물으니 "미세먼지수준은 보통입니다"라고 답함. 아이들이 “기가지니가 우리목소리를 일일이 다 인식해서 각자 애청하는 프로그램을 알아서 도와줬으면 좋겠다” 고 함	Positive
D20	사람이 많을 땐 음성인식을 사용하지 않게 됨. 시간이 너무 오래 걸임 리모콘보다 빠를 수는 없음	Neutral
	기가지니를 이용해서 티비를 시청하고 동시에 핸드폰을 이용해서 물건을 주문함	Neutral
	TV는 음성인식으로 켜고 리모컨으로 예능, 영화, 드라마 검색 후 '나 혼자 산다' 시청함. 중간에 음성인식으로 시간을 물었는데 답하니까 편하다고 느낌	Positive
D21	엠비씨를 틀라고 했으나 못틀고 자유라고 응답.	Negative
	음성 검색으로 유투브 채널 검색(역시 빠름). 방탄소년단 뮤직비디오 DNA를 검색했으나 잘 못알아 들음. 원치 않는 연관 채널이 많이 나옴	Neutral
D22	기가 지니를 이용해 분유 수유를 하며 좋아하는 드라마 신청	Neutral
	음성인식으로 보라카이 여행후기 검색. 유투브로 연결. 스마트폰으로 함께 검색하면서 시청함	Neutral
	리모컨을 기가지니 검색으로 찾은 뒤 TV다시보기에서 '백종원의 골목식당' 시청함. 편리하긴 한데 음성인식이 좀 더 친근하게 대화하듯 쓰여지면 좋겠다는 생각을 함	Positive
D23	사람이 많을 땐 리모컨만 사용하게 됨	Neutral
	출근준비를 하면서 뉴스를 듣다가 음성인식기능으로 날씨를 확인하였음	Neutral
	음성인식으로 리모컨 찾음. 갑자기 음성인식 오류로 인식도 안되고 종료도 안되고 이전버튼을 누르니 겨우 꺼짐. 자주 발생한다	Negative
	기가지니에게 ‘코미디빅리그’를 틀어달라고 해서 시청함. 내가 원하는 장면만 짤방으로 볼 수 있으면 좋겠다고 생각함	Neutral

D24	기가지니한테 뉴스 틀어 달라고 해서 보면서 아이 이유식 먹이면서 뉴스 시청	Neutral
	리모컨으로 TV 켜고 음성인식으로 사남매쇼 검색해서 시청.	Neutral
D25	채널 이동, 볼륨 조절에 음성인식 기능 사용함	Neutral
	다시보기로 아는 형님을 보다가 9시반경 음성인식으로 tvn을 틀어 드라마를 보았다	Neutral
	아들이 일어나서 기가지니에게 채널을 바꿔달라고 하면서 만화를 시청함. 기가지니를 계속 불렀는데 소리가 켜져 있으면 잘 안 들리는 듯 하다. 그래서 소리를 질러도 답이 없어 가족들이 모두 답답하다고 말함.	Negative
	Ai스피커로 볼륨조절 시도했으나 잘 인식이 되지 않아 여러번 시도.	Negative
	기가지니로 리모컨 찾음. 이 기능이 제일 편리하다	Positive
	음성검색으로 유투브 검색-> DIY 연관채널 시청 중 '휴대폰케이스 만들기'검색을 아이들이 시도했으나 검색이 잘 되지 않아 다시 '만들기'로 검색함	Neutral
D26	집안 정리를 하며 음악을 들었다. (지니에게 틀어달라고 함)	Neutral
	저녁식사 후에 가족들이 다같이 모여 기가지니에게 신서유기를 찾아달라고 했다. 신기하게도 금세 잘 찾아 주어 가족들이 모두 기뻐했다. 기가지니에게 처음부터 끝까지 시키고 싶지만 말을 못 알아 들을 때가 많아 포기했다.	Positive
	"기가지니 리모컨 찾아줘"는 이제 일상이 됨. "기가지니 고독한 미식가 틀어줘'로 검색한 뒤 시청함. 음성인식은 상당히 편한 기능이라는 걸 점점 더 느낌	Positive
D27	볼륨 조절을 잘 알아듣지 못함	Negative
	저녁식사 준비를 하면서 음성인식으로 TV를 켰다 리모컨이 없어도 TV를 켤 수 있는 점이 제일 맘에 든다 다른 기능은 명령어를 고민하느니 그냥 리모컨을 돌리는게 편하다 말을 좀 더 잘 알아듣고 데이터 연동이 스마트폰만큼 잘 되면 쓸모는 있을 거 같다	Positive
	날씨 정보를 더 상세히 들으려고 인공지능에 날씨를 물어 봄	Neutral
	이제 가장 많이 쓰는 기능은 "기가지니, 리모컨 찾아줘"임.	Neutral
D28	아기가 잘 때는 딱히 음성인식으로 할 이유가 없다. 리모컨이 편함	Neutral
	음성인식 기능이 좀 더 개선되면 좋을 것 같다. tv를 켜고 끄는건 자연스러운데 채널을 빠르게 검색하거나 다른 기능을 쓸 땐 여전히 음성인식이 불편하다. “tv켜줘” “JTBC” 정도로는 아주 잘 쓰고 있다	Positive
	기가지니에게 날씨 물어보고 TV 켜달라고 해서 봄	Neutral
	지니를 쓰면 리모콘 쓸때 같이 리모콘을 들었다 놨다 할 필요가 없으니까 집안일 하면서 채널 돌리고 하는데 자주 사용했는데 오늘 조금 불편해도 최대로 활용했다고 생각했다. 어머니와 수미네반찬을 보고 따라해 보자며 수미네 반찬을 틀고 요리를 했는데 요리를 하면 손이 리모콘을 잡기 부적합해 질때가 많아서 일시정지, 재생 하고 말하며 유용하게 썼다. 하지만 조금 딜레이가 생긴다는 점이 아쉬웠다.	Positive

	백종원의 골목식당' 시청하면서 친척들이 방영된 장소에 가고 싶다고 함. "기가지니, 저 장소 검색해 줘" 하면 장소를 찾아주는 기능도 있으면 좋겠음	Neutral

국문초록

이것은 인터랙티브 TV 사용자 경험 (user experience, UX)에 관한 연구이다. Human-Computer Interaction (HCI) 분야에서 시스템 성능이나 사용성은 사용자 만족에 영향을 미치는 주요 요소로 평가되어 왔다. 시스템이 특정 분야에서 사용되거나 사용자 – 시스템 간 상호작용 범위가 제한적인 때는 이러한 접근 방식이 매우 효과적이었다. 하지만 기술의 발달로 이제 우리 주변에는 수 많은 기술이 존재하게 되었으며, 사용자는 이러한 기술을 자유롭게 선택하고 사용할 수 있게 되었다. 이는 기술의 유용성이나 사용성 이상의 가치를 필요로 함을 의미한다. 실제로 어떤 기술은 금새 사라지기도 하고 어떤 기술은 오랫동안 살아남거나 우리 삶을 더욱 풍성하게 만드는데 기여하기도 한다. 기술의 불확실성과 복잡성 때문에, 이제 산업계는 기술 그 자체보다는 사용자의 동기나 경험에 더 초점을 맞추고 있다. 사용자 경험은 성능이나 사용성뿐 아니라 재미나 즐거움 등 겉으로 들어나지 않는 감성적 품질을 포함하는 포괄적 경험을 의미한다.

TV는 우리 일상에서 가장 쉽게 접할 수 있는 엔터테인먼트 미디어 중 하나이다. 기술의 발달에 따라 새로운 기능들과 결합하면서 TV 역시 인터랙티브 제품으로 진화하고 있다. 하지만 실제로 인터랙티브 TV의 다양한 기능들이 적극적으로 활용되지 못하고 있으며 이는 사용이 불편하거나 필요 없기 때문으로 나타나고 있다. 이에 본 연구는 (1) 인터랙티브 TV 사용 경험에 영향을 미치는 주요 경험 품질 요소를 알아보고, (2) 인터랙티브 TV 사용 경험 확장을 위한 사용자 인터페이스를 비교 평가하며, (3) 인터랙티브 TV UX 모델을 구성하여 사용자 경험과 사용 만족도, 그리고 새로운 기술의 지속적인 사용 의사 간의 인과관계 검증을 연구 목표로 정하고 실증적 연구를 진행하고자 한다. 이를 위해 가장 먼저 사용자 경험에 대한 선행 연구와 UX 모델을 검토하고 연구 방법론에 대해 살펴 보았다. 첫 번째 연구 목표인 인터랙티브 TV 사용 경험에 영향을 주는 품질 요소를 찾기 위해 TV를 시청하는 사용자 행동 상황을 관찰

평가하였다. TV 시청자의 주요 행동 유형으로 목적 지향적 상황과 특별한 목적을 갖지 않고 과정 그 자체에 의미를 두는 상황을 도출하고 각각의 상황에 대한 사용자 경험을 평가하였다. 두 번째 연구 목적인 인터페이스 대안 평가를 위해서는 기존의 TV 조작 방식인 리모컨과 인터랙티브 TV 사용에 적합한 대안 중 하나로 떠오르고 있는 대화형 에이전트 (conversational agent, CA) 간 비교 평가를 실시하였다. 사용자는 두 가지 방식을 각각 이용해서 TV 시청과 관련된 과제들을 수행하고 이러한 경험에 대해 스스로 평가하였다. 동시에 사용자가 상호작용 중에서 느끼는 기분이나 감정을 객관적으로 측정하기 위한 생체 반응 측정도 실시하였다. 이와 별도로 매일 TV 시청 경험을 기록하는 필드 테스트도 진행하였다. 필드 테스트는 20가구를 대상으로 4주간 일상에서 리모컨 또는 CA를 조작하면서 TV 시청 경험을 기록하도록 하고 매 주 한 번씩 사용자 만족도와 지속 사용 의사를 측정하여 사용 경험에 따라 만족도와 사용 의사의 변화가 나타나는지 관찰하였다. 마지막으로 인터랙티브 TV UX 구조방정식 모형을 통해 사용자 경험이 사용 만족도에 미치는 품질 요소와 사용자 경험과 사용 만족도, 지속 사용 의사 간의 인과관계를 검증하기 위한 확증적 요인 분석을 실시하였다.

일련의 연구 결과는 다음과 같다: 사용자 경험은 단순히 시간을 보내기 위함이 아닌, 구체적인 시청 목적을 갖고 있을 때 더 긍정적으로 나타났으며 이러한 사용 경험에 영향을 주는 요소는 “simple,” “controllable,” “innovative,” impressive,” and “original” 등이었다. 경험 품질은 사용성 차원보다 감성 차원에서 더 강하게 나타났다. 사용자 인터페이스 비교 실험에서 사용자 경험에 영향을 미치는 품질 요소는 “manageable,” “predictable,” “stylish,” “innovative” 등이었다. 생체 반응 측정 결과, 리모컨과 CA 조작 과정에서 사용자의 인지적 노력과 스트레스가 서로 다르게 나타남도 확인하였다. Skin Conductance (SC) 측정 결과, CA 상호작용에서는 주로 스트레스가 높음 → 낮음의 변화를 보였으나 반대로 리모컨 상호작용에서는 낮음 → 변화 없음 또는 낮음 → 높음의 변화를 보이는 경우가 많았다. 이는 사용자가 CA 에게 음성 명령을 내리는 시작 단계에서 인지적인 부담을 느낄 가능성을 의미한다. 반대로 리모컨의 경우, 조작하는

방식은 인지적 부담은 크지 않으나 메뉴를 확인하거나 여러 번 버튼을 누르는 과정에서 스트레스가 증가할 수 있음을 발견할 수 있었다. 이러한 결과에 따르면, CA와 리모컨의 역할이 서로 다르며 따라서 CA는 리모컨의 대체수단보다는 보완역할에 적합한 것으로 보인다. 이런 점에서 현재 리모컨 사용에 최적화되어 있는 TV 인터페이스는 사용자 경험 확장을 위해 CA와 리모컨 사용을 모두 고려한 인터페이스로 전환이 필요하다.

실 사용자를 대상으로 한 필드 테스트에서는 평가 전 측정 기준 대비, 4주간 사용 만족도가 꾸준히 증가하는 것으로 나타났다. 사용자의 지속사용의사는 2, 3주까지 오히려 감소하다 4주에 다시 증가하는 형태를 보였다. 이에 대해 사용자 다이어리를 살펴 본 결과, 사용 초기에는 음성 인식률 오류 등으로 인한 불만이 매우 높았으나 반복적인 사용을 통해 CA 품질 수준을 인지하고 받아들이는 단계를 거치는 것을 알 수 있었다. 마지막으로 구조방정식 모형을 통해 인터랙티브 TV UX 모델 프레임워크를 구성하고 사용자 경험 품질을 확인하였다. 구조방정식 모형을 통해 도출된 경험 품질 요소는 “predictable,” “stylish,” “premium,” “challenging” 등으로 나타났다. 모든 연구에서 공통적으로 도출된 interactive TV 사용 경험 품질 요소는 “manageable,” “innovative,” “captivating,” 그리고 “novel”이다. 확인적 요인분석 결과, 긍정적 사용 경험 품질이 사용 만족도에, 그리고 사용 만족도가 지속적인 사용 의사에 각각 인과관계를 나타냈다. 인터랙티브 TV 사용 만족도와 지속 사용 의사에 영향을 미치는 경험 품질은 사용성 품질 차원보다 감성 품질 차원에서 더 크게 기여했다. CA 역시 기술적으로는 미완성 단계이지만 감성 품질 차원에서 사용자에게 더 어필하는 것으로 나타나, 향후 대화형 방식이 인간–컴퓨터 상호작용의 새로운 패러다임이 될 수 있음을 예측하게 한다. 본 연구를 통해 사용자의 경험과 경험에 기반한 만족도가 새로운 기술의 intention to use를 예측하는 의미 있는 척도가 될 수 있을 뿐만 아니라 기술의 성공 여부에는 긍정적인 사용자 경험이 필수적이라는 점을 확인하였다. 복합적 방법을 적용하여 다각도로 인터랙티브 TV 사용자 경험을 살펴봄으로써 사용자 경험 연구 결과의 신뢰성을 높일 수 있다는 가능성도 엿보았다. 이번 연구가 TV 뿐만 아니라 CA 기술을 적용하는 다른 인터랙티브 시스템 개발에서 더 긍정적인 사용자

경험 제공을 위한 참고 자료로 활용될 수 있기를 기대한다.

Keywords: interactive TV, user experience, everyday products, conversational agent, hedonic quality, ergonomic quality

Student number: 2011-31008