

# Evaluating User Experience Using Physiological Data from Conversational Interfaces for Interactive TV

Seungyeon Lee<sup>1</sup>, Chan Hyeok Yun<sup>1</sup>, Gee Won Shin<sup>1</sup>, Soo Yeon Kim<sup>1</sup>, Myung Hwan Yun<sup>1</sup>

<sup>1</sup>Department of Industrial Engineering, Seoul National University

## ABSTRACT

**Objective:** This study aims to investigate TV viewers' user experience (UX) of a conversational agent (CA) that assisted interactions while watching TV. **Background:** TV is evolving into an interactive product as technology advances, but the TV interface has not changed much. This can limit the UX of interactive TV viewers. Therefore, it is necessary to identify how the use of voice-interactive interface affects the TV users in order to have a better interactive TV experience. **Method:** To comparatively analyze the UX with the CA interface and the traditional remote control unit (RCU) interface, physiological measurements (skin conductance; SC) for objective data as well as self-report questionnaires were evaluated. **Results:** Physiological measurement results showed that SC differed between CA and RCU interactions. Most SC data were high at the beginning of the CA interaction but decreased over time. In RCU sessions, SC maintained a constant value or increased. Consequently, CA had a greater effect on the overall attractiveness while having more emotional appeal. **Conclusion:** Subjective results showed that CA provided a more positive TV experience regarding emotional aspects. **Application:** Physiological measurement can provide a more detailed and rich interpretation of the emotional changes and cognitive efforts during human-TV interaction, leading to better understanding of the UX mediated by interactive agent.

**Keywords:** user experience, conversational agent, voice recognition, physiological measurements, ergonomic quality

**Corresponding author:** Myung Hwan Yun (mhy@snu.ac.kr)

## 1. Introduction

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 quickly. However, despite the technological changes, the TV usage pattern of viewers has not changed much. People do not use interactive functions in TV, because

it was uncomfortable or not necessary (Darnell, 2007). Unlike TV's evolution, the main interface is still a remote control unit (RCU), which may not be suitable for interactive environments such as text input and contents navigation. Conversational agent (CA), based on the voice recognition technology, appeared in the early 2000s, but not frequently used because of low voice recognition rate and lack of database. 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. Especially in Korea, CA is often combined with everyday products such as TV rather than being used alone (Consumer Insight, 2017). According to Bushnell (2018), Murnane (2018), and Reeves (2017) who reviewed Amazon Alexa, CA does not yet appear to be at full interactive communication stage. Contrary to these claims, Yuan (2016), Hill, Ford and Farreras (2015), and Purington et al. (2017) argued that users replied that they felt more familiar in CA interactions. In terms of interactive TV experience, it is necessary to know whether the CA interface contributes to the interactive TV experience by playing a more positive role in the interactive TV experience. In this study, we apply the physiological measures and subjective user evaluation to find out what different experience the user has in interactive TV interaction by comparing two interfaces, RCU and CA. Among physiological measures, skin conductance (SC) measures are regarded as an index of arousal, anxiety, or stress (Hudlicka, 2011). 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. Mandryk et al. (2006) and Yao et al. (2014) used physiological measures to evaluate user experience as well as users self-report questionnaires. Another important reason is that it is possible to grasp specific emotional changes during the interaction process.

## 2. Method

All 40 graduated students (20 males and 20 females) participated in this experiment. The average age of the participants was 29.1 years. Two experiment were performed to operate the TV using each interface, CA and RCU. Before attending the experiment, sensors were attached to the fingers of participants to detect SC data in interaction. In order to balance the experiments, the first 20 participants performed the first experiment with CA, and then used RCU. From the 21<sup>st</sup> participant, RCU interaction was conducted first and then CA was used. After each session, participants evaluated their own experiences when they used interfaces in watching TV. Six tasks which are related to TV control were carried out: channel zapping (T4, T9), menu navigation (T6), VOD control (T8a, T8b), and content search (T10).

## 3. Results

### 3.1 The results of physiological measures

Repeated measurement analyses were carried out to examine if there were differences in SC measurement values for each task performed. For all the tasks, the data of 13 seconds were extracted from the data that recorded for 20 seconds by excluding the dummy data for 7 seconds at the front part. The SC result showed significant differences partially or overall in the 5 tasks among the total of 6 tasks (see Table 1).

**Table 1. The result of repeated measures of SC data in the CA and RCU interactions**

|                 |              | M.S. | F     | Sig.    |
|-----------------|--------------|------|-------|---------|
| T4              | Interface(I) | .055 | .072  | .790    |
|                 | Time(T)      | .023 | .719  | .535    |
|                 | I*T          | .047 | 1.079 | .354    |
| T6              | Interface(I) | 2.72 | 4.78  | .036*   |
|                 | Time(T)      | .077 | 3.41  | .030*   |
|                 | I*T          | .144 | 7.08  | .000*** |
| T8 <sup>a</sup> | Interface(I) | .27  | .43   | .519    |
|                 | Time(T)      | .06  | 1.65  | .078    |
|                 | I*T          | .06  | 3.05  | .000*** |
| T8 <sup>b</sup> | Interface(I) | .25  | .30   | .586    |
|                 | Time(T)      | .06  | 1.97  | .026*   |
|                 | I*T          | .03  | 1.67  | .072    |
| T9              | Interface(I) | .69  | .90   | .349    |
|                 | Time(T)      | .12  | 3.95  | .000*** |
|                 | I*T          | .01  | .55   | .878    |
| T10             | Interface(I) | .32  | .36   | .555    |
|                 | Time(T)      | .04  | .78   | .671    |
|                 | I*T          | .30  | 5.25  | .000*** |

\*\*\*  $p$  value < .001, \*\*  $p$  value < .01, \*  $p$  value < .05

Note. (T4, T9: channel zapping), (T6: menu navigation), (T8a, T8b: VOD control), (T10: content search)

T6 showed a significant difference between CA and RCU in variables of interface ( $F(1, 34) = 4.78, p < .05$ ), time ( $F(12, 408) = 3.41, p < .05$ ), and interface and time ( $F(12, 408) = 7.08, p < .001$ ). There were significant differences in the scores for CA and RCU in T8<sup>b</sup> ( $F(12, 348) = 1.97, p < .05$ ), and T9 ( $F(12, 408) = 3.95, p < .001$ ) over time. T8<sup>a</sup> ( $F(12, 372) = 3.05, p < .001$ ) and T10 ( $F(12, 396) = 5.25, p < .001$ ) showed a significant differences by interface and time. T4 showed no significant difference between the two interfaces in terms of interface, time, and interface and time.

After comparing the results between RCU and CA, commonalities for each graph type were also analyzed (see Figure 1). Most of the CA interaction results show high initial stress and then low, regardless of the level of difficulty. RCU showed no change in stress (T6, T8a) or initially in a low level and increased stress over time (T4, T8b, T10). 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.

Figure 1 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. 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.

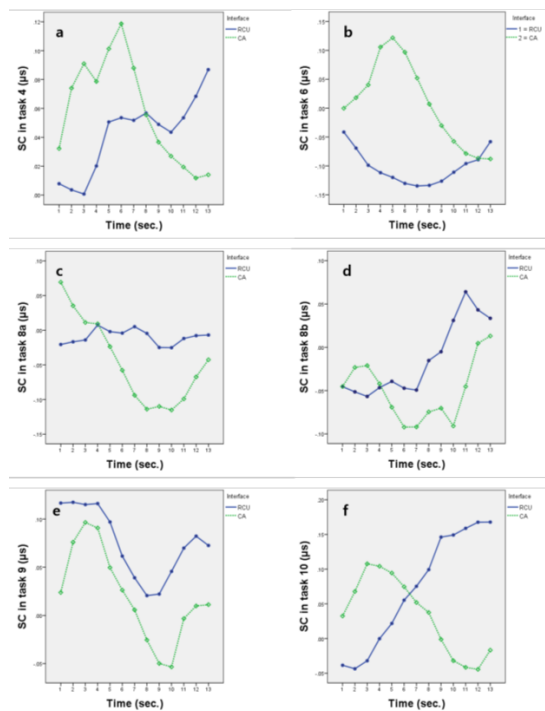
### 3.2 The results of user experience satisfaction

**Table 1. The results of user satisfaction in interaction**

| UX satisfaction | RCU mean | CA mean | t     | sig. (2-tailed) |
|-----------------|----------|---------|-------|-----------------|
| Pleasant        | -.80     | .24     | -3.07 | .004**          |
| Attractive      | -.05     | .63     | -3.29 | .002**          |
| Likeable        | .02      | .68     | -2.40 | .021*           |
| Inviting        | .22      | .93     | -2.22 | .032*           |
| Good            | .39      | .76     | -1.42 | .165            |
| Appealing       | .10      | .37     | -1.17 | .248            |
| Motivating      | -.20     | .27     | -1.93 | .060            |

\*\*\*  $p$  value < .001, \*\*  $p$  value < .01, \*  $p$  value < .05

In this study, the semantic differential scale was used to evaluate the user satisfaction (Hassenzahl et al., 2010). Participants self-evaluated questions that asks their own feelings after each session. The question consists of seven items related to user satisfaction, with scales ranging from -3 to +3. A two-tailed paired t-test was conducted to compare the difference between the CA interaction and the RCU interaction for significance (see Table 2). As shown in Table 2, CA was evaluated more positive than RCU in all results. "Pleasant" ( $t(40) = -3.066$ ,  $p < .01$ ), "attractive" ( $t(40) = -3.285$ ,  $p < .01$ ), "likeable" ( $t(40) = -2.402$ ,  $p < .05$ ), and "inviting" ( $t(40) = -2.220$ ,  $p < .05$ ) were significantly larger than in the CA interaction. "Good,"



“appealing,” and “motivating” were not statistically significant difference between two interfaces. In summary, participants were a considerably more pleasant and felt attractive and inviting experience in the CA interaction.

## 4. Conclusion

The purpose of this study is to investigate how the combination of interactive agents and TV changes the user experience using objective and subjective methods. Physiological measurements are useful for collecting objective data and identifying changes in user feelings through the analysis of physiological data during interaction. Watching TV is similar, but the emotional responses of the user during the CA interaction and the RCU interaction were opposite. The result of physiological measurements is meaningful in that it can provide a more detailed and rich interpretation of the emotions and cognitive efforts during human-TV interaction. This result can also be used as a reference for TV viewing scenarios or for the interaction design of CA equipped products. It is also interesting that participants evaluate their interaction with the CA as more attractive than the RCU. The emotional quality of the experience contributed even more to the overall attractiveness and satisfaction. It was found that CA is more appealing and more positively affecting users' emotions. However, it is essential to maintain a practical perspective in that CA is adopted for products that are used primarily in daily life, and that usability is one of the most important attributes of such products. The results of physiological measurements are considered useful enough to yield specific data about the process of user experience. We hope this research will help provide meaningful results in our research on interactive technology and user experience mediated by interactive agents.

## References

- 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-assist-ant-amazon-alexa-face-off.html>
- 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](https://www.consumerinsight.co.kr/voc_view.aspx?no=2924&id=ins02_list&PageNo=1&schFlag=0)
- Damell, 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.
- Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, interactive products – Facets of user experience. *Interacting With Computers*, 22 (5), 353-362.
- 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.
- Hudlicka, E. (2011). *Affective Computing: Theory, methods, and applications*. CRC Press, Boca Raton.
- 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.
- Mumane, K. (2018, May 3). Dumb And Dumber: Comparing Alexa, Siri, Cortana And The Google Assistant. *Forbes*. Retrieved from <https://www.forbes.com/sites/kevinmumane/2018/05/03/dumb-and-dumber-comparing-alexa-siri-cortana-and-the-google-assistant/#14f1c96436e7>
- Purinton, 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.
- 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-world-through-words-alone-how-voice>
- 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.
- 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-conversation-al-design/> University of Georgia, Points of Pride, *University of Georgia*, <http://www.uga.edu/profile/pride.html> (retrieved October 21, 2009).