My Smart TV Agent: Designing Smart TV Persona for Linguistic UX

Seyeon Lee, Jiwon Moon, Hokyoung Im, Chung-Kon Shi, and Bong Gwan Jun

Graduate School of Culture Technology, KAIST
291 Daehak-ro(373-1 Guseong-dong), Yuseong-gu, Daejeon 305-701, South Korea
{birdkite,moon.jiwon,imagine21c,bonggwan,chungkon}@kaist.ac.kr

Abstract. Linguistic UX design for Smart TV has been creating much heap as a means of new TV control. Since human voice displays powerful social presence, the issue with defining a Smart TV agent that interacts with users has a big impact in users' satisfaction. The purpose of this study is to analyze the linguistic patterns in vocal commands of TV users and to suggest underlying personas of Smart TV agent that appears when users interact with a Smart TV. First, we analyzed most common TV viewing situations and the patterns of users' behavior through a survey. Then, we collected 867 vocal data through a cultural probe method in which 10 families, each representing a typical type of TV viewers, by asking them to record what they would like to say to the TV while watching it for about a week. We suggest 6 different type of Smart TV personas, such as expert, assistant, colleague, slave, machine and pet, based on the relationship that the user and TV exhibited. With the collected vocal data, we analyzed the participants' speech pattern and style to examine which type of Smart TV persona was most prevalent. As a result, there were slight difference in types that emerged according different functions of Smart TV and we found that the assistant type appeared most frequently followed by the colleague type.

Keywords: Affective communication, Smart TV agent, Linguistic UX, Voice command.

1 Introduction

We no longer expect our TV to just change the channel and adjust the volume for us. TVs can now perform a lot more functions with the birth of Smart TV which is capable of bringing the internet to the TV screen [1]. However, it is difficult to encourage users to actively engage with them just because there are more functions to operate since TV is the longest standing Lean Back media in people's mind. That is why there is a steady effort to find the most appropriate way to encourage users to engage in active control of TV utilizing a touch remote control, a motion control, smartphones, motion capture and vocal command etc.

The most attractive attribute of vocal command technology is that users can easily pick it up without particular training. However, at the moment, it fails to meet the standard people expect to get from the experience. Of course, the level of communication

with high level of cognitive ability to understand context only existed between human to human communications. Therefore, users will experience something different when commanding TV using vocal communication compared to using a remote control or a key board.

For linguistic UX design, there is a strong emphasis in affective approach based on the understanding of human psychology and behavior. The machine that is interacting with the user is now perceived as a machine that can understand the user's intent and communicate with a personality [2]. Then, which persona of Smart TV most likely to satisfy the user and overcome the barrier of uncanny valley? What type of relationship should this persona (or the agent) should represent? The goal of this research is to look at different types of implied listener that seems to appear in the communication process between the user and Smart TV. Also, we will suggest appropriate persona of the TV agent that the users will be most likely to be satisfied with.

2   User - Smart TV Relationship

According to Chatman, S.B., there is a certain narrative-communication situation model that describes the author and the reader in story and discourse [3]. In case of users' effort to vocally command Smart TV, there is an implied speaker and a listener representing the user and the Smart TV. It can be said that this implied listener who implements what user wants as the agent inside of a Smart TV. Then, how can we define the relationship between the user and the Smart TV agent?
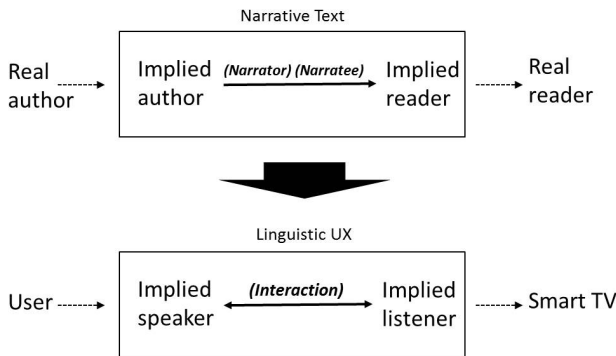
Fig. 1.　"Narrative-Communication Situation Model" applied in linguistic UX for Smart TV

The human-computer relationship can be described with three examples: One-up vs. One-down vs. One-across. First, by One-down they argued that the computer is just another tool and that it should just follow the command of the user. Secondly, One-up implies that the computer is a master with high knowledge that is able to carry out difficult things the user cannot otherwise. Finally, One-across suggests a team-work between the computer and the user come together equally as a team [4]. In this case, the researchers found out that the user and the computer felt a sense of

connectedness [5]. Such relationship between a user and a computer can be defined by the language the users use when interacting with a computer. Therefore, through a linguistic analysis of TV viewers' situation, we can examine how the users of Smart TV define their relationship with the machine.

Personalities of Smart TV agent are also important for building a relationship with the users, and it can also affect user satisfaction. Among five important factors of defining personalities that psychologists suggested [6], the two most important categories of media personalities are categories like dominance/submissiveness and friendliness/unfriendliness [4] [7] [8]. Users will show different types of speech style according to not only their own personality but also in accordance to the personality of the Smart TV agent they interact with. This is relevant to the suggested relationship between the Smart TV and the users. For example, if a Smart TV shows dominant trait over the user, the user will respond with more respectful tone of speech and if submissiveness is the set personality of the machine, then users will talk down or use direct commands. When the Smart TV displays friendliness, users will use intimate style of speech and not when the machine shows unfriendliness. Therefore, we can figure out the personality of the Smart TV agent when we look into the words and speech style of the TV users.

In addition to specify the persona of a Smart TV, factors such as gender and age must be examined. It is typical of users to respond in regard to their own stereotype that differs by gender and age of the counterpart [9]. This is why it is important to find which gender and age would be most appropriate for a Smart TV agent in order to satisfy the expectation of the users. In this line of examination, we need to look at whether it is more adapting to change the personality of a Smart TV agent according the TV contents and the users' behavior.

3   Method

In order to analyze how users verbally interact with a TV, a two-step examination was conducted. First, in order to find general TV viewing patterns according to viewers' age and gender and to select typical TV viewers, we conducted a survey. Based on the results of this survey and through an interview, we selected participants for our cultural probe experiment to collect vocal command data.

3.1   Survey

An online survey was released aimed at gathering data on TV viewers' behavior and the choice of content inherent to the moment of watching TV based on age and gender. For 2 weeks, 173 participants (m: 76, f: 97) each described 5 most common TV viewing behaviors at home. We were able to collect 722 data on TV viewing situation and observed different type of TV viewing behavior based on age and gender.

We divided TV viewers into 9 groups according to their age and explored which type of TV programs were most frequently watched by each group, with whom, and what they usually did when they were watching TV. As for the children under 10,

kids programs were most frequently watched with their parents accompanying them. For teenagers, they usually played games or spent time engaging with social networking services while watching TV. Many in their 20's and 30's said that their TV was on during resting moments at home, before going to bed and while eating as well. Gender divided the preference of TV programs among young people with sports/game and entertaining TV contents preferred by males and drama and entertainment programs for women. As for men in their 40's, many watched the news. Women in their 50's watched drama with an exceptionally high preference. Based on these results and interviews, we recruited participants who showed similar typical TV viewing patterns and conducted a Cultural Probe Research.

3.2   Cultural Probe Research

In order to extract natural voice commands from participants as if they were interacting with a Smart TV in real life, cultural probe methodology was utilized to collect such data. Cultural probe methodology is a useful way to collect data by allowing participants to record their own behavior, situations and their ideas freely in everyday lives [10]. We were able to recruit appropriate candidates for our Cultural Probe research based on collected information of the TV viewers' characteristics of age, gender and behavior from our survey. Eleven selected participants (m: 4, f: 7/age under 10: 1, 10's: 1, 20-30's: 6, 40-50's: 2, over 60: 1) were given a kit with a voice recorder and a journal to record the family's TV viewing environment. Specifically, they were asked to write down the names of the TV programs, who were watching with them and what they were doing while watching TV. Also, the families were asked to record whatever they wish to say to the TV they were watching. From this observation we were able to collect 867 vocal data from 10 families.

4   Results

4.1   Smart TV Function and Four Sentence Types

First, we categorized the vocal command data into 6 different types of Smart TV functions: Basic Function, Search Function, Recommend Function, Additional Function, and Social talk. Basic function refers to functions that can be performed by the TV remote control such as adjusting the TV volume and changing the channel etc. Search function allows the user to explore information. Vocal commands for asking for the weather and a specific TV program schedule etc. would be such case. Recommend function enables the user to get a recommendation from the TV. Vocal data such as "Turn on some music program" and "Put on a documentary about pets" can be classified as commands for such function. Additional function implements the user's request that is different with the previously mentioned functions and is new to the TV experience. Vocal commands like "Capture this scene", "Show me from the beginning", and "Skip advertisements" are some of the examples. Social talk refers to innocent utterances of the user with no intention of commanding the TV.

Second, with the vocal data, we examined the frequent use of the 4 types of speeches such as the declarative, imperative, interrogative, exclamatory type. Then we cross-examined it with the 6 Smart TV functions.

Table 1. Number of vocal data in different function and sentence types

	Declarative	Imperative	Interrogative	Exclamatory
Basic Function	11	143	1	0
Search Function	9	154	169	0
Recommend Function	7	55	24	0
Additional Function	43	203	19	0
Social Talk	13	1	9	6
Total	83	556	222	6

Overall, the use of the imperative sentences were dominating in all functions, however, there were some differences in each function. For the basic function, the imperative sentence covered 92.3% and other types of sentences were rare to find. Interrogative sentences took up 50.9% of the search function request, implying that users generally requested to get more information from the Smart TV while watching something. Imperative and interrogative sentences each scored 64% and 27.9% for the recommend function. As for the additional function, imperative sentences highly used with 76.6%, however, declarative sentence type followed that number with 16.2%. Social talk was rare to detect, however, declarative and interrogative sentences were most frequent and it was the only function to attract exclamatory sentences.

Although we were able to find out which function had the most frequent request when users were watching TV and which style of speech they were using, such results were not sufficient in finding out which Smart TV persona the users were engaging with. Therefore, in the next step of our research, we analyzed the users' speech style in detail and drew out 6 types of TV agent personas according to how the users were verbally implying their relationship with the Smart TV.

4.2   Six Types of Smart TV Persona Analysis

Users' vocal commands were divided into 6 different styles of speeches by two factors, respect and intimacy. First, when there is lack of intimacy and high respect, people use honorifics. In Korean language, there are certain grammar rules to be applied when speaking with an elder person or those who you respect making the style of speech significantly different. Such unique characteristic allowed us to pick out vocal data for the first speech type. Second, when there is strong intimacy and respect, people use an expression like "let's" while using informal language and more intimate words. The third type of speech style is when both intimacy and respect are average, typical vocal commands would include the word "please" and the expression "can you…" translated in English. When the level of intimacy and respect are both low, users use expressions that have no respect for the listener, a typical vocal command would sound like "(just) do it." The fifth speech type appears when there is strong

intimacy but little respect. There is no expectation of the other to respond but the speaker uses friendly form of speech. Finally, the last type of speech shows when there is almost no sense of intimacy and respect for the other. The speaker addresses the other as a non-human being and commands as if they would type in a command on a computer.

Table 2. Six type of Smart TV agent categorized by speech style

| Type	Intima-cy	Respect	Speech style	Example (Translated in English)
Expert	Low	Very High	Honorific language	· It would be great if you could pro-vide translated subtitle. · Could you please show me the shopping channel? · I would be happy if you could add the share function.
Colleague	High	High	Let's~ I hope~ Do you Know~?	· Let's change the music. · I hope the TV turned off when I fall asleep. · Do you know where I can buy it?
Assistant	Neutral	Neutral	Please, do~ Will you~? Can you~?	· Please change the channel. · Let me know the location of that café. · Will you tell me the price of those shoes?
Slave	Low	Low	(just) do~	· Change to channel 5. · Turn up the volume. · Record this song.
Pet	High	Low	Monologue /Utter	· I will came back after eat some food. · The ghost is typing Morse code. It's scary! · I hope to go there. · It looks delicious!
Machine	Very Low	Very Low	Word /Command (not a sen-tence)	· KBS (name of the channel). · Alarm. · Forward.

By categorizing the vocal data into 6 different types, we can glimpse how the users are defining their relationship with the Smart TV. We named each type according to their features as the following: Expert, Colleague, Assistant, Slave, Pet, and Machine. We then checked the frequency of these types in our vocal data.
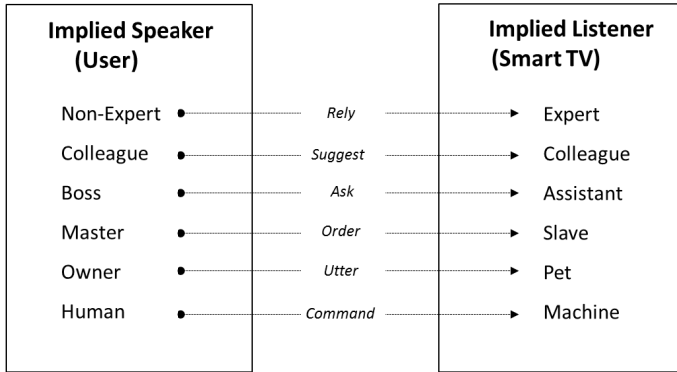
Fig. 2. User-Smart TV relationship and implied verbal manners

Out of 867 vocal commands that we collected, the most prevalent command was "please do (this)" with 391 cases, which is about 45% of the whole data. And the Assistant type with such wording took up 51%. The second most frequent type was the Colleague type with 29%. Generally, expressions like "let's do…", "I would like you to…", and "I want to do…" that were indirectly commanding the TV and questioning about a TV program were included in this case. Therefore, many users were interacting with an underlying TV persona by showing low level of intimacy and respect but feeling comfortable with an Assistant type who can help them or a Colleague type who is an equal being to them.

In addition, we found that there were only a handful of cases indicating a Machine type which has no humanistic feature; and there was hardly a case for the disrespected Slave type as well. This may indicate that when people use their own voice as an input device to control Smart TV, they are treating it as a human being that has a sense of respect. The Expert type took up 7% of the data; however, we discovered that it was one particular participant who has been mainly using honorific type of speech. Therefore, it is difficult to generalize using formal type of speech to a Smart TV.
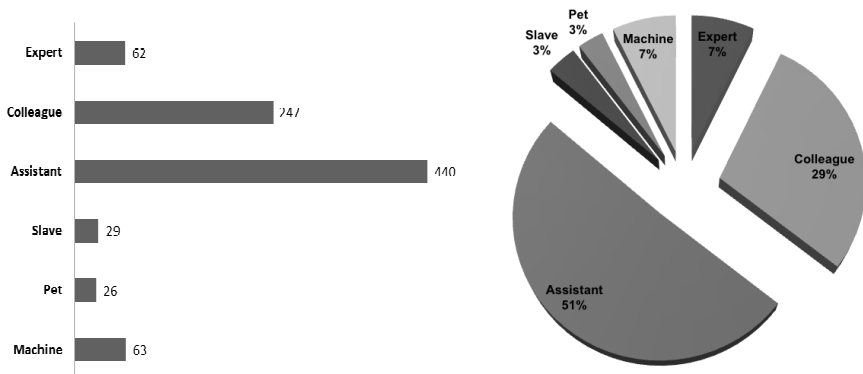
Fig. 3. Number and percentage of vocal data appearing six different types of agent

4.3   Smart TV Functions and Six Types of Smart TV Personas

In this step, we wanted to find different types of Smart TV personas for the TV functions we categorized in the previous section. As a result, the Assistant type was dominant in most functions. The Assistant type appeared most frequently with 72%, especially in the basic function.

On the other hand, for the search function, the Colleague type appeared most frequently with 50% and followed by 41% of Assistant type. As we analyzed in the previous section, interrogative sentences are used mostly for the search function. Since many participants used causal tone of asking for information or to control, which is relevant to the Colleague type, this type dominates almost the half of all data. Even though these kinds of expressions are more closed to the Colleague type, they also can be used to refer the Assistant type.

Table 3. Number of vocal data in different functions and six types of agent

Function	Expert	Colleague	Assistant	Slave	Machine	Pet
Basic Function	7	5	111	14	17	1
Search Function	10	167	137	0	18	0
Recommend Function	10	26	45	3	1	1
Additional Function	35	40	146	12	28	4
Social Talk	0	9	1	0	0	19

Even if the number of vocal data for the "Pet" type was very small, 66% of them appeared as social talk function. In this experiment, participants would say something to the TV, but the TV couldn't provide a feedback. Thus, such dialogue becomes a monologue similar to talking to a pet. If a Smart TV is developed to a certain degree so that they can give a witty answer to users' sporadic expressions, we can assume that the level of intimacy and respect would increase, making it easier for the Colleague (friend) type to appear more often.

5   Discussion and Conclusion

This study is the first step of developing appropriate Smart TV personas from the linguistic UX point of view. From this study, we were able to find a certain linguistic tendency of users' voice commands and several implied personas in users' general situation of TV viewing.

Our research showed that many users didn't regard Smart TV as just a machine or a tool. Instead, they seemed to interact with it as if it were a living thing. In addition, most of data represented that users regarded Smart TV as an assistant, who has a lower status than the user, or as a colleague who has the same level of status with the user. Furthermore, we were able to find a possibility that the relationship between the

user and Smart TV could be developed to provide more equal sense of the relation in the future if advanced Smart TV provides more valuable experience to the users. Natural language processing in the electronic equipment such as computers, mobile phones, and Smart TVs has not been developed to the level of reproducing human-like natural linguistic communication, and users are not yet accustomed to the linguistic UX. Since users have a low expectation of the Smart TV's ability to have a natural conversation, we can assume why the percentage of "Social talk" among the whole functions was very low.

The limitation of our study is that the collected vocal data does not reflect the real life interaction between a Smart TV and a user; the participant recorded their vocal command without hearing the response of the TV. We considered using an existing Smart TV which is applied in linguistic UX researches for collecting vocal data, but since the role of Smart TV is yet ambiguous, we wanted participants to freely imagine the role of the TV and record their dialogues. In the next step, we can analyze various dialogues and feedbacks from the users when they experience a two-way communication with Smart TV.

In addition, we conducted the experiment with only Korean participants and all the collected vocal data was in Korean language. Since we analyzed the vocal data based on the linguistic rules of Korean language, it may be difficult to generalize the findings when translated into other languages. Lastly, this research did not cover to suggest the preferred age, gender, and detailed personalities of a Smart TV agent depending on different groups of viewers. Therefore, for future work, we will look at the preferences of Smart TV personas depending on different age and gender groups. Furthermore, we will develop TV agent characters that reflect users' preferences and evaluate the users' level of satisfaction and affection with them.

References

1. Shin, D.H., Hwang, Y., Choo, H.: Smart TV: Are they really smart in interacting with people? Understanding the interactivity of Korean Smart TV. Behaviour & Information Technology 32(2), 156–172 (2013)
2. Nass, C., Steuer, J., Tauber, E.R.: Computers are social actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 72–78. ACM (1994)
3. Chatman, S.B.: Story and discourse: Narrative structure in fiction and film. Cornell University Press (1980)
4. Reeves, B., Nass, C.: The Media equation: How people treat computers, television, and new media. Cambridge University Press (1997)
5. Nass, C., Fogg, B.J., Moon, Y.: Can computers be teammates? International Journal of Human-Computer Studies 45(6), 669–678 (1996)
6. John, O.P., Donahue, E.M., Kentle, R.: The "Big Five. Factor Taxonomy: Dimensions of Personality in the Natural Language and in Questionnaires". In: Pervin, L.A., John, O.P. (eds.) Handbook of Personality: Theory and Research, pp. 66–100 (1990)
7. McCrae, R.R., Costa, P.T.: The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. Journal of Personality and Social Psychology 56(4), 586 (1989)

8. Nass, C., Moon, Y., Fogg, B.J., Reeves, B., Dryer, C.: Can computer personalities be human personalities? In: Conference Companion on Human Factors in Computing Systems, pp. 228–229. ACM (1995)
9. Nass, C., Moon, Y., Green, N.: Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. Journal of Applied Social Psychology 27(10), 864–876 (1997)
10. Gaver, B., Dunne, T., Pacenti, E.: Design: Cultural probes. Interactions 6(1), 21–29 (1999)