Voices of the Internet of Things: An Exploration of Multiple Voice Effects in Smart Homes

Yohan Moon[1], Ki Joon Kim[2], and Dong-Hee Shin[3(✉)]

[1] Department of Interaction Science, Sungkyunkwan University, Seoul, South Korea
ttattang@skku.edu
[2] Department of Media and Communication, City University of Hong Kong, Hong Kong, China
stand4good@gmail.com
[3] School of Media and Communication, Chung-Ang University, Seoul, South Korea
dshin1030@cau.ac.kr

Abstract. Based on the Computers Are Social Actors (CASA) paradigm, this study investigates an effect of media specialization by number of voice in smart-home environment where many smart devices are controlled by voice user interface (VUI). Result from a between-subjects experiment (N = 50) examines that there are interaction effects between users personality and number of voice on social attraction and trust toward media technology which are critical in human-computer interactions. In this experiment, extrovert users feel a stronger feeling of social attraction and trust when there is one identical voice from several smart devices. On the other hand, introvert users feel a stronger feeling of social attraction and trust when different smart devices make respective voices. These results provide a strong evidence for human's automatic social response to smart devices which have a voice, a strong anthropomorphic cue. Finally, we discuss on implications for future VUI setting, according to user's personality.

Keywords: Voice user interface · CASA · Natural user interface · Number of voice · User personality

1 Introduction

How do users perceive smart devices with different voices? Would users prefer single, identical voice or multiple different voices in smart devices? In a recently released movie "Her" (2013), a relationship between a lonely men and operating system which talks is illustrated. The scientific fiction gives an implication that people might possibly build a relationship with an agent which has a voice.

As natural user interface (NUI) has been an ultimate goal of user interface design [24, 25], Voice User Interface (VUI) has been pointed as promising way of user interface [16]. There has been a plethora of research on how users perceive a voice from a computer or a robot [12, 13, 19, 23]. However, how do users perceive multiple voices of different smart devices? This question has become increasingly important as the era of Internet of Things (IoT) emerges. As the IoT is filling our routine with various multiple voices of

The original version of this chapter was revised: The affiliation of the third author was corrected. The erratum to this chapter is available at 10.1007/978-3-319-39862-4_46

N. Streitz and P. Markopoulos (Eds.): DAPI 2016, LNCS 9749, pp. 270–278, 2016.
DOI: 10.1007/978-3-319-39862-4_25

smart devices, the investigation on how user perceives multiple voices from respective smart devices is getting important for socially meaningful user experience. Therefore, this study examines whether the number of voice (one voice x three voice) of smart device influences on user's feeling of social attractiveness, and trust toward media technology according to user personality (introvert x extrovert) in smart home environment.

1.1 Voice User Interface as Natural User Interface

A Voice User Interface is what a human interacts with in communication with a spoken language application [4]. As an ideal way for interaction with computer, voice user interaction has been pointed out. Because voice interfaces are regarded "more natural", compared to other types of interfaces (e.g., keyboard, mouse, touch screen), human-computer interaction by voice was inspired and motivated. Voice interfaces have a "look and feel" which is similar to human-human communications. The given assumption is that the more natural interface is, the more people would perceive and be accustomed to the interface easily and effectively. Providing "more natural" interface enables system to make use of skills and expectations that people has evolved through routine experiences for effective and expeditious communication [8, 15].

Science fictions have shown VUI when controlling machine by just talking only a short time ago. With the advance of technology, VUI have become more prevalent and people are practically making use of the usefulness that eyes-free and hand-free interface provided in numerous situations. Voice-based virtual privacy assistant agent embedded in various types of computers, ranging from laptop to mobile phone to smart devices, can execute numerous tasks. They transform voice to text, interpret requests from users, and seem to comprehend and execute what users ask, and even interact conversations based on online data and its own accumulated database.

In the past, there was also insistence that a successful human-machine interaction, similar to successful human-human interaction, was goal to execute the task effectively from the human's perspective. However, human-computer voice-based interactions of those days did not yet meet the accuracy, reliability, richness, or complexity developed in most human-human interactions. The deficiency was due to imperfect technology of voice recognition [8]. However, nowadays a successful level of recognition has been developed so that operating system like Siri or Now can comprehend what we say. Not only the reasons of its natural interaction way referred above, but this positively also affects various facets of user experience such as multitasking when doing some other things with our two hands, the ease of use when the task include too many screen touch. Especially helping not to select too many screen menus may helps users get tired of decision of choice by reducing moment of choice [2, 26].

Nowadays, voice-based virtual privacy assistant agent such as Apple Siri, Google now, Microsoft Cortana and Amazon Alexa can interpret natural speech. The advent of these voice-based agents had drawn an attention to the voice control system [29]. Moreover, Voice user interface is a type of way of interface which enables the users to send emails, schedule an appointment, turn on music, and more [7]. As the popularity increase, the voice agent is everywhere, functioning as ambient computing. From the mobile phone to smart phone to smart watch, to smart home environment such as smart

speaker or smart TV, it has been embedded to multiple applications slowly and widely, as mostly privacy intelligent assistant agent.

2 Theoretical Approach

2.1 The Media Equation Theory: Do People Equate Smart Devices with Genuine Social Actor?

The meaning of ''media equation'' is that ''individuals' interactions with computers, TV, and new media are fundamentally social and natural, just like interaction in real life'' [27]. Because human have perceived that all the objects were real and only human had its own human-like shapes and characteristics like language, emotion, personality, rapid interaction, and so on. Human brains had evolved to treat anything that looks to be real as real and anything that seems to have anthropomorphic characteristics as real human. Therefore, when people face the any kind of media such as TV or computer, the limitation of perceiving everything at its outward value and responding to virtual action of media as if they were real occurs because of evolutionary reason. In this way, the media equation occurs [14]. Especially on human responses to computers, Nass, Steuer, and Tauber presented Computers Are Social Actors (CASA) paradigm [23], which means that individuals unconsciously apply social rules as if they were interacting with real human beings when interacting with computers that show anthropomorphic cues. The base of this research paradigm is that, if computers or machine people confront with have anthropomorphic cue, individuals automatically respond to them socially and do not perceive that they are not real human.

2.2 Voice as Social Attribution

Voice, rather than shape, has been pointed as key factor of social attributions toward computers [23, 28]. The important of voice as social cue has been argued in the previous researches on the HCI field in the aspect of CASA paradigm. With interactivity and filling of roles, words of output has been considered as important primary cues, which is held by humans, human-like characteristics. These kinds of cues seem to automatically induce schemata related with human-human interaction, without the psychological construction of relevant human [20].

One of the five experiments conducted to build the CASA paradigm demonstrated that "people respond to different voices as if they were distinct social actors, and react to the same voice as if it was the same social actor, regardless of whether the different voice was on the same or different computer" [23]. This result shows the implication that people perceive social actors as many as they perceive the number of voice.

2.3 Theories of Attraction

There are two equally convincing social rules related to personality in and HCI literatures and interpersonal interaction—similarity attraction and complementary attraction. The similarity attraction rules insist that people are more stick to people who are similar to themselves, and prefer to interact with them. In accordance with this rule, perceived

similarity, which means a degree of what we believe something is similar to ours, is sufficient to make us attracted to others [5]. Demographics, ethnicity, political attitudes, and personality are examples of making us believe that we are similar to others. The complementary attraction rule insists that people are apt to be attracted people who have opposite personality characteristics, so that their personalities make balance, complementary situation [6, 28]. However, there are just few studies in the comparison to researches supporting similarity attraction rule.

2.4 Extended Concept of Media Specialization

An abundance of research on specialization has been investigated in multiple contexts, demonstrating that technology which is assigned a specific role or area is perceived as specialist [9–11, 22]. In those previous studies, it is found that specialized technology, which has one specific label and functions the particular role, induces users to trust and prefer more than generalized technology, which has simultaneously two or no specified label, though those two technology perform same. For example, in the experiment conducted by Nass [23], participants watched the same news and entertainment materials on television sets. There were two different conditions of TV labels, the 'News TV' and 'Entertainment TV' (i.e., specialist) or 'News/Entertainment TV' (i.e., generalist). Participant who watched News and Entertainment on the specialist TV evaluated the contents higher and preferred them more than who watched the same contents on the generalist TV.

Based on categorization theory, label has been regarded as important signifier which triggers a set of related social category-based perceptions [1]. Initial impressions of an object are constructed primarily based on social categories brought by salient cue [9], such as label. In previous researches, media specialization was mainly

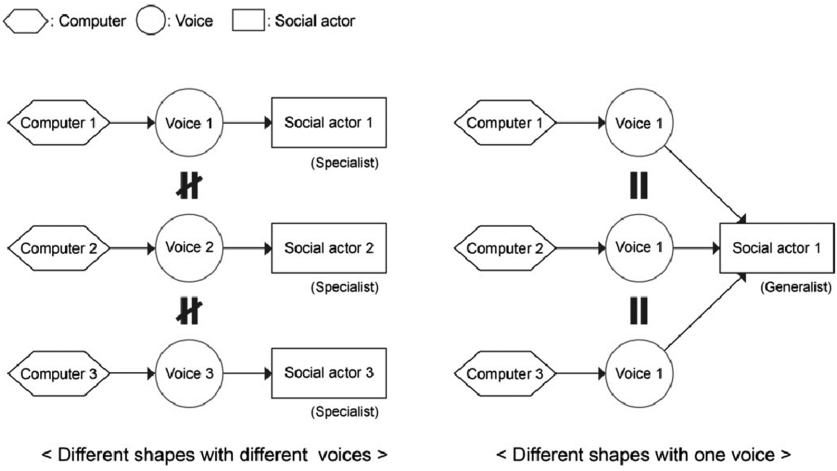

Fig. 1. Concept model proposed on relationship between computer, voice and social actor in the aspect of CASA.

made by labeling, such as attaching role name tag on the device. It has been thought to be a strong and effective way of specialization.

However, would it be possible to make specialization of the media with their voices? According to the concept of specialist and generalist, people tend to apply social response that individuals more trust and prefer the people who take only one role than people who takes simultaneously two or more roles when interacting with technology. Voice is distinctive social cue which makes individuals discern several social agents [18]. According to researches such as CASA paradigm and the media equation theory [23, 27], each one voice is perceived as one social actor. From the extension of interpretation of those researches, next hypothesis is inferred (Fig. 1).

Hypothesis 1. Users will have a greater feeling of trust to smart home environment when devices which have respective functions have their respective different voices than when those devices have one identical voice.

2.5 User Personality

User personality has been investigated in related HCI researches [12, 21]. For example, Nass and his colleague conducted experiment to observe when similarity attraction appears, dividing participants introvert and extrovert according to conditions [19]. Extroverts have a tendency that they are more social and outgoing than introverts [17]. According to similarity attraction rule [3, 21], the next hypotheses are inferred:

Hypothesis 2. Extroverts will be more attracted to smart device with multiple voices than one identical voice.

Hypothesis 3. Introverts will be more attracted to smart device with single voice than multiple voices.

Taking these hypotheses into consideration, the next research question is followed.

RQ: What is the relationship between user's personality and number of voice of smart home environment in user's social response to smart home environment?

3 Method

3.1 Experimental Design and Participants

A 2 (number of voice: one vs. three) X 2 (user personality: introvert vs. extrovert) between-subjects experiment was conducted. A total of 50 undergraduate students (18 males, 32 females) were recruited for the experiment through an online registration page. A web-based Wiggins [30] personality test was administered to participants prior to main experiment. Participants were randomly assigned to the conditions, with both gender and personality approximately balanced across the conditions.

3.2 Materials

Three voices were recorded from eight different women with specific guide on four voice parameters (speech rate, volume level, frequency, and a pitch range) [14] to keep voice's

personality neutral. Among the eight recorded voices, only three voices which were rated as the most neutral and clearly different from the other were chosen by 10 people interview. To make respectively different voices, Text To Speech (TTS) or computer-generated speech were not used. Three smart devices (TV, speaker, and lamp) are set in a mirror room with small Bluetooth speakers hidden respectively behind the respective smart devices.

3.3 Procedure

Participants were told that they were going to test a new voice recognition service developed for smart home system. Two basic tasks for each device were assigned such as changing volume of smart speaker or changing channel of smart TV in a two-way mirror room with only verbal control. Participants were given the scenario guiding the order to control, and were informed to finish tasks one by one. The experimenter controlled the smart devices and Bluetooth speakers behind the mirror room. Before the task starts, participants were informed to talk to the devices in a way they say to Siri, the voice-based virtual agent, and to proceed in the order of the scenario. The scenario included not very specific such as script but the order of usage among the three devices. In a two-way mirror room, set like a furnished home with sofa, participants are seated and manipulated the devices by talking to them. Participants were randomly assigned to conditions. Participants could hear the recorded feedback message after every verbal control. For example, when controlling speaker, if the participant said "turn on any Jazz music", then the smart speaker responded, "Okay, here is the Jazz Music" and the Jazz music prepared was played. After a while, if participant told that they didn't like the song which was being played and they wanted another kind of music, speaker said "what kind of music do you like?" and played the genre the subject responded and said "How is this music?". In case of additional unexpected request of participants, a bunch of materials such as music and responses were recorded and prepared in advance.

3.4 Measures

After completing the tasks with three devices (TV, speaker, lamp) as guided, each participant answered a questionnaire. Items measuring social attraction (Cronbach's $\alpha = 0.90$), trust toward media technology ($\alpha = 0.82$) were adopted from [14, 10], respectively. All the variables were measured using 10-point Likert scale ranging from 1 = "not at all" to 10 = "very much so."

4 Result

Two-way analyses of variance (ANOVAs) were conducted to analyze the effect of the number of voice and type of user personality on the dependent variables. The results revealed no significant main effects of the both independent variables.

However, significant interactions between the number of voice and type of user personality predicting social attraction, $F(1, 46) = 12.44$, $p < .01$, and trust toward media

technology, $F(1, 46) = 12.04$, $p < .01$, were observed (see Fig. 2). These results indicated that the extroverts experienced greater feelings of social attraction and trust toward media technology when they were exposed to a single, rather than multiple voice, whereas the introverts felt stronger feelings of social attraction and trust when interacted with multiple, rather than a single voices.

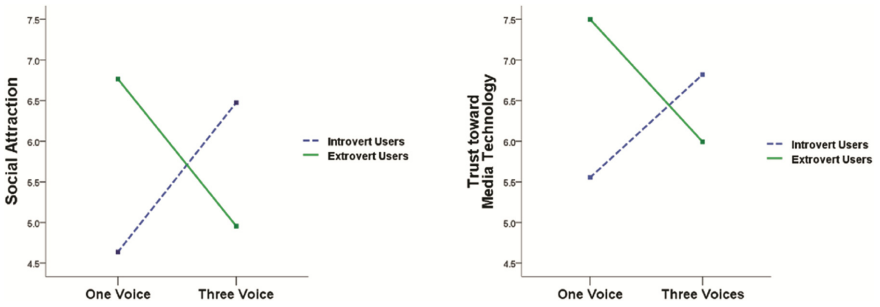

Fig. 2. Interaction effects between number of voice and users personality

5 Discussion

Results show that there are significant relationship between number of voice of smart devices and user's personality, while all hypotheses are not supported. It is assumed that there is another aspect of similarity rule which plays an important role in leading opposite result. One of the features of extroverts is that they are more talkative than introverts [12, 17]. When every smart product which is pervasive in smart home environment has the only one same identical voice, the situation may make extroverts perceive that there is the only agent who take care of all the devices and the agent embedded in smart device is also talkative because the only voice appears consistently in every respective device. Inversely, when every smart product in smart home environment has the several different respective voices, the situation may make extroverts feel that they are talking very shortly with different people. Therefore, they may perceive that they are staying with agents who are not talkative.

It is important to craft user interface with detail consideration. A carefully manipulated user interface can exceed numerous limits of current technology to accomplish a successful user experience, even when the technology functions imperfectly [8]. In the smart environment, natural user interface is getting important and this study may pose a critical design implication of voice user interface as natural user interface.

References

1. Ashforth, B., Humphrey, R.: The ubiquity and potency of labeling in organizations. Organ. Sci. **8**, 43–58 (1997)
2. Baumeister, R.: Ego depletion and self-control failure: an energy model of the self's executive function. Self and Identity. **1**, 129–136 (2002)

3. Byrne, D., Griffitt, W., Stefaniak, D.: Attraction and similarity of personality characteristics. J. Pers. Soc. Psychol. **5**, 82–90 (1967)
4. Cohen, M., Giangola, J., Balogh, J.: Voice User Interface Design. Addison-Wesley, Boston (2004)
5. Infante, D., Rancer, A., Womack, D.: Building Communication Theory. Waveland Press, Prospect Heights (1990)
6. Isbister, K., Nass, C.: Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. Int. J. Hum. Comput. Stud. **53**, 251–267 (2000)
7. Jung, S., Lee, K.M., Biocca, F.: Voice control system and multiplatform use: specialist vs. generalist? In: Yamamoto, S., Abbott, A.A. (eds.) HIMI 2015. LNCS, vol. 9172, pp. 607–616. Springer, Heidelberg (2015). doi:10.1007/978-3-319-20612-7_57
8. Kamm, C.: User interfaces for voice applications. Proc. Natl. Acad. Sci. **92**(22), 10031–10037 (1995)
9. Shin, D.: User experience in social commerce: in friends we trust. Behav. Inf. Technol. **32**(1), 52–67 (2013)
10. Koh, Y., Sundar, S.: Effects of specialization in computers, web sites, and web agents on e-commerce trust. Int. J. Hum. Comput. Stud. **68**, 899–912 (2010)
11. Koh, Y., Sundar, S.: Heuristic versus systematic processing of specialist versus generalist sources in online media. Hum. Commun. Res. **36**, 103–124 (2010)
12. Lee, K.M., Nass, C.: Designing social presence of social actors in human computer interaction. In: Proceedings of the Conference on Human Factors in Computing Systems - CHI 2003, pp. 289–296 (2003)
13. Shin, D., Choo, H.: Modeling the acceptance of socially interactive robotics: social presence in human-robot interaction. Interact. Stud. **12**(3), 430–460 (2011)
14. Lee, K., Peng, W., Jin, S., Yan, C.: Can robots manifest personality?: an empirical test of personality recognition, social responses, and social presence in human-robot interaction. J. Commun. **56**, 754–772 (2006)
15. Leiser, R.: Improving natural language and speech interfaces by the use of metalinguistic phenomena. Appl. Ergon. **20**, 168–173 (1989)
16. Lim, Y.: Disappearing interfaces. Interactions **19**, 36 (2012)
17. McCrae, R., Costa, P., McCrae, R.: Personality in Adulthood. Guilford Press, New York (1990)
18. Nass, C., Gong, L.: Speech interfaces from an evolutionary perspective. Commun. ACM **43**, 36–43 (2000)
19. Nass, C., Lee, K.: Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency-attraction. J. Exp. Psychol. Appl. **7**, 171–181 (2001)
20. Shin, D.: Defining sociability and social presence in social TV. Comput. Hum. Behav. **29**(3), 939–947 (2013)
21. Nass, C., Moon, Y., Fogg, B., Reeves, B., Dryer, D.: Can computer personalities be human personalities? Int. J. Hum. Comput. Stud. **43**, 223–239 (1995)
22. Nass, C., Reeves, B., Leshner, G.: Technology and roles: a tale of two TVs. J. Commun. **46**, 121–128 (1996)
23. Nass, C., Steuer, J., Tauber, E.: Computers are social actors. In: Paper Presented to CHI 1994 Conference of the ACM/SIGCHI, Boston, MA, USA (1994)
24. Negroponte, N.: Being Digital. Knopf, New York (1995)
25. Norman, D.A.: The Design of Everyday Things. Basic Books, New York (2002)
26. Pocheptsova, A., Amir, O., Dhar, R., Baumeister, R.: Deciding without resources: resource depletion and choice in context. J. Mark. Res. **46**, 344–355 (2009)

27. Reeves, B., Nass, C.: The media equation. CSLI Publications, Stanford (1996)
28. Sullivan, H.: The Interpersonal Theory of Psychiatry. Norton, New York (1953)
29. Shin, D.: User value design for cloud courseware system. Behav. Inf. Technol. **34**(5), 506–519 (2015)
30. Wiggins, J.: A psychological taxonomy of trait-descriptive terms: the interpersonal domain. J. Pers. Soc. Psychol. **37**, 395–412 (1979)