Voice Control System and Multiplatform Use: Specialist Vs. Generalist?

Soyoung Jung^{1(\Box)}, Kwan Min Lee², and Frank Biocca^{1,3}

¹ Syracuse University, S.I. Newhouse School of Public Communication, 201 University Ave., Syracuse, NY 13202, USA sjungOl@syr.edu

² University of Southern California, Annenberg School for Communication and Journalism, Los Angeles, CA 90007, USA

³ Interaction Science, SungKyunKwan University, Sungkyunkwan-Ro, Jongno-Gu, Seoul, Korea

Abstract. How do participants respond to and prefer either a common voice agent that follows them across platforms or a community of specialist agents connected to specific hardware platform. This study examines users' gender as it relates to preference for voice-control system (VCS). The participants experiences a VCS agent that followed them across platforms such as from a smart pone to a personal cloud computing (PCC) or smart TV. One group met a different specialist agents that were tied to different devices. This study pits the effect of specialization by interface against the principle of consistency design principle with regards to voice agents. However, we found a strong gender effect, females preferred a single, generalist female agent across platforms while males tended to prefer different, specialist female agents embodying different platform.

Keywords: Voice control system \cdot Voice agent \cdot Multi-platform \cdot Clouding \cdot Specialist vs. Generalist \cdot CASA HCI

1 Introduction

How do users perceived their relationships with voice agents that embody and represent interaction with a device? In a recently released movie "Her" (2013), a lonely man develops a relationship with a talking operating system, named "Samantha." This scientific fiction dramatizes the well, established research evidence that users have conscious and unconscious social responses to synthesized voices and agents (Nass, Steuer, Tauber, & Reeder, 1993; Nass, Steuer, & Tauber, 1994; Reeves & Nass, 1996; Nass & Lee, 2001).

But what kind of interaction would be best for users and voice agents in a multiplatform or multi-device environment? Would users prefer and respond more strongly to a single, consistent voice agent that follows users across platforms? Or would they prefer and trust specialist agents who "specialize" in their interaction with a particular platform. This study compares two types of voice agent control systems—specialized agent vs. consistent (generalist) agent—while crossing the multiplatform environment of user devices, focusing on how users accept and evaluate the two different types of interactive agents. The study explores both theoretical and practical implications for voice agent studies. Also, the theoretical approach can enhance the study of affective or cognitive source orientation with the gender effect.

1.1 Voice Control System (VCS)

Popularity. Voice interaction has been seen an ideal for interacting with technologies. The revolutionary speech interpretation and response interface "Siri," an artificial-intelligent, "assistant" (Trautschold, Ritchie, & Mazo, 2011), has drawn world wide attention to this kind of interaction (Watson 2012). More generally voice control systems with natural speech-based user interfaces, like Siri, Google now, and Microsoft Cortana, have gained wide circulation in interface interactions.

Define applications. The VCS is a type of voice control application (app.), which allow the users use their voice to send messages, schedule meetings, place phone calls, and more. Ask VCS to do things just by talking the way users talk. VCS process voice to text, interpret what users are requesting, appear to understand what the users mean, execute tasks, and even provide conversational interaction on accumulated database. The accumulated database help VCS to assist the users.

Siri and system. The current systems such as Siri have achieved a successful threshold of recognition: they enhance users' experience of the multi-tasking ability of the interfaces, the ease of use and practicality of voice, as well as with the apparent "personality" of the voice interface. According to this increasing popularity, voice agent has been augmenting several application areas including texting, scheduling meetings, searching, emailing but also other applications areas (Pirani et al. 2014; Jaramillo et al. 2014).

Different types.

Users on mobile. Furthermore, the popularity of smartphone has led to an escalated use of VCS, because of the ability to control devices with less hand interaction (Jaramillo et al., 2014). Moreover, mobility of smartphone encourages people to use more VCS. By using mobile VCS, users can send execute commands and messages while they are on the move. Pirani and et al argue that the use of mobile devices have generally encouraged users to use VCS in a variety of settings (2014).

1.2 Agents in Multiplatform and Cloud Computing Environments

Cross platform. Owing to the increasing worldwide popularity of voice control system (VCS) in smartphones, there is a diffusion and adoption in the use of voice intelligent personal agent (IPA) across different interface platforms. So users experience agents as they navigate across platforms from phone to their TV or laptop computer.

Different apps. For example internet-connected smart television (TV) provides interactive features, including all kinds of applications available on the Web.

Why multeplatform. With increased functionalities, however, interactive elements interfere with screen-obscuring content, which may harm the quality of the watching

experience when using the system (Berglund and Johansson 2004). Berglund et al. (2004) suggested new user interface and service solutions with speech and dialogue for interactive TV navigation. These has been pursued in several systems. Voice agent interactions can be seen in automobiles and other platforms.

Cloud computing. Cloud computing also creates and environment relevant to agents. Users are interacting with their data across many platforms. Private cloud computing (PCC) storage is a service used to store synchronous data. This allows end users to access their data from multiple platforms, "anytime, anywhere" from iCloud, Skydrive, and Evernote (Armbrust 2010). Cloud-computer users save their photos, documents, and email on the virtual storage of their cloud service so that they can use the synchronous data wirelessly while they switch among platforms, such as a laptop, desktop, tablet personal computer (PC), or even a smart TV. Moreover, Ye and Huang (2011) have introduced a framework for a cloud-based smart home.

Agents that share some "familiarity" with user data across platforms make consistency of interaction of potential value in this environment. The agent moves across platforms like the data. Given the utility of such cross-platform and VCS app, this study will examine to what extent the attitude towards a general cross platform agent clashes with the specialized voice agent and how this can affect user perception of the agent, the device, and the information provides.

2 Theoretical Approach

2.1 Computers Are Social Actor

There is a research base on user responses to computers as social actors. Nass and Lee (2001) examine the computer-synthesized speech and its personalities. According to "Computers are Social Actors" or "CASA" theory (Nass, Steuer, Tauber, & Reeder, 1993; Nass, Steuer, & Tauber, 1994; Reeves & Nass, 1996), when humans interact with computers, users perceive the computers as social actors regardless of the interfaces. Therefore, researchers who subscribe to CASA theory, tend to apply social psychological theories and principles to human-computer interaction (HCI).

2.2 Consistency-Attraction and Specialization

There is an existing body of research in social psychology and HCI regarding how people perceive "generalists" and "specialist" be it people or interfaces.

Nass and Lee (2001), examined how similarity and consistency-attraction works for computer-synthesized speech in HCI studies. Consistency-attraction is the idea that people prefer consistent characteristics while they interact because if people have interaction with inconsistent characteristics, this creates a cognitive load and makes it hard to predict what will happen when they engage with others (Field 1994; Fiske and Taylor 1991; Thomas and Johnstone Thomas and Johnston 1981). When users perceived a social presence from a computer, users were preferred to a consistent computer synthesized voice and text (Nass and Lee 2001). People with a tendency towards

anthropomorphism are more likely to prefer same (generalist) voice agent than those with a less tendency towards anthropomorphism. It is assumed individuals who tend to anthromorphize perceive the voice agent as a social actor.

Media Specialization. At the same time, there is perceived social value in "specialization." Findings in previous study show that the devices that have "specialized" functions are seen as superior to devices that have "generalized" functions be they monitors, sensors, and other devices (Nass and Moon 2000; Reeves and Nass 1996; Sundar and Nass 2000; Sundar and Nass 2001). As specialized computer devices become more common, this value for specialization is more frequently applied to media technologies and becomes more important and salient feature. Moreover, Reeves and Nass found that only labeling as a "specialized" device or introduced as specialized media has significant effects on users in their work on the media equation (1996). The attributions relating to specialization and generalization have affect on perceivers' attitude and evaluation toward technology and media (Leshner et al. Brewer 1998; Nass et al. 1996, 1994; Nass and Steuer 1993). For example, Nass and et al. examined the attribution of specialization and generalization in TV set in 'Technology and Role: a tale of two TV' and a specialist TV set is evaluated as having a better screen and better content deliver TV set than a generalist TV set, even though they are both same physically. Not only the TV set but Web agent, web site and computer were also examined to testify the credibility of specialist effect (Koh and Sundar 2010).

Media Integration. However, Reeves and Nass (1996) doubted that specialization would cause conflicts to occur when integrating across different functions in the media equation. Also, they worried that using a "single" media appliance could "create a sense of commonality across the different functions" (p. 152). These two significant design principles—specialization and consistency—would be competing principles. This study's purpose is to investigate the effects of specialization and consistency and to determine which principle is more effective as voice agent interfaces cross platforms.

2.3 Research Question

In the crossing-multiplatform situation (between smartphones and smart TV), would a specialist (varied agent; Siri and Sori) or generalist (consistent agent; Siri and Siri) voice agent be perceived as more effective in users' evaluation of the social perception of the agent, the agent's usability, and their attitudes towards to the quality of the interface and the viewing experience?

Individual differences: Given that social interaction is affected by gender patterns, would the gender of the users and two different agents affect the evaluation of the agent, the usability tests of the voice control system, and the viewing experience?

Hypothesis 1. When users experience a voice agent for the first time, the gender of the user lead to differences in the social perception of the agents.

Hypothesis 2: When users experience agents across platforms, they will perceive specialist agents as more social, usable, and effective than a generalist agent.

Hypothesis 3. The gender of the user (male or female) will respond differently to specialist or generalist agents.

3 Method

3.1 Experiment Design

For this experiment a between-subject design was used with two factors: Type of agent with two levels, consistent or specialized voice agent, and the second factor, user gender with two levels, male vs. female.

3.2 Participants

A total of 36 participants were recruited via an online bulletin boards for payment. Participants blocked by gender, 18 males and 18 females, and randomly assigned to one of two voice agent conditions.

3.3 Materials

Three interfaces were created (See Fig. 1). First, we used a common phone interface featuring a voice agent called, Siri. We created two interactive, smart television interfaces, one featuring the same agents as the phone, Siri, the other using a new voice agent, Sori. Both Sori' and Siri were played in female Korean synthesized voices, and based on translated scripts from the American version of Siri. Sori has a similar conversational repertoire as Siri, except that Sori introduces herself as a specialist.



Fig. 1. Interfaces by different type of voice agent.

Procedure. Users entered the lab and were instructed that they would evaluate a new interface. All participants were introduced to a smart phone using the voice agent, Siri.

The participants were escorted to a simulated living room with a one-way mirror, and then, they were guided to the sofa. The participants received an iPhone- featured device that featured the voice agent, Siri, but the device was controlled via a Wizard of Oz technique, meaning that when subjects interact with computer system they believe that the system works autonomously, but the researcher behind a the one way mirror remotely manipulates functions.

Brief instructions on how to control the voice- recognition system were given to participants, along with four tasks—listening to music, watching news, scheduling, and calling. After completing the four tasks with the iPhone and Siri, each participant went out to answer a questionnaire using a computer. During the first phase of the experiment, all participants experienced the Korean version of Siri.

After filling out the questionnaire, half of the male or female participants were randomly assigned to interact with the same agent, the Korean Siri, for the voice-control system for the smart TV while the other half of the participants were assigned to new agent Sori for the voice-control system for the smart TV.

The participants faced a smart TV and controlled the smart TV with voice agent to finish tasks, which was manipulated by a researcher from outside of the simulated room using the same Wizard of Oz technique used in the first phase of the study. As in the first phase of the study instructions on how to use VCS were given, and the four tasks were explained. In the both conditions the agents greeted the participants, but the specialist agent condition, Sori added that she was a specialist for smart TVs. After completing the four tasks with different contents on the smart TV, each participant went out to answer a questionnaire on a computer.

After completing all of the tasks and the questionnaire, all of the participants were debriefed regarding the purpose of the study, thanked for their participation, and dismissed.

3.4 Measures

In order to observe the evaluation of agents both on the smartphone and on the smart TV, we measured **social perception** of the agent using: social presence (Cronbach's $\alpha = .82$, and Cronbach's $\alpha = .83$, social attraction (Cronbach's $\alpha = .83$, and Cronbach's $\alpha = .87$), and perception of likability (Cronbach's $\alpha = .91$ and Cronbach's $\alpha = .87$) of the agent, i.e., the social presence was measured by five statements, such as "I focused on the interaction with the voice control system," and "I felt that I was really communicating with the agent." In order to evaluate the agent on 10-point scales, 15 questions were asked (Fig. 2).

The assessment of the **overall usability** of the voice control systems was measured using five sub-scales: usefulness (Cronbach's $\alpha = .93$, and Cronbach's $\alpha = .91$), ease of use (Cronbach's $\alpha = .93$ and Cronbach's $\alpha = .93$), ease of learning (Cronbach's $\alpha = .91$, Cronbach's $\alpha = .89$), satisfaction (Cronbach's $\alpha = .91$ and Cronbach's $\alpha = .84$), The sub-scales of 19 questions about usability were asked using 10-point scales: "It helps me be more effective," "It is simple to use,". Attitudes towards the experience and product were measured using content evaluation was inquired about using three questions, i.e., "I am satisfied with the content" (Cronbach's $\alpha = .92$, and Cronbach's $\alpha = .91$); viewing experience, 10-point semantic differential scales were used with the following adjectives: enjoyed, excited, and had fun (Cronbach's $\alpha = .90$ and Cronbach's $\alpha = .80$), and buying intention (Cronbach's $\alpha = .86$, and Cronbach's $\alpha = .84$).

4 Results

4.1 Results of Experiment 1

The first phase of the experiment was a control condition. An independent t-test was executed to examine the effects of the participants' gender with regard to the voice-control system in the smartphone control setting across all ten measures in the area of social perception, usability, and attitudes towards the experience and product. There were no significant differences across for gender for any of the 10 measures¹. So the groups were similar in their responses to voice agents in first phase of the study.

A two-way, between-groups ANOVA was conducted to explore the impact of gender on the type of agent, generalist or specialized, experienced when viewing the smart television No significant differences were found in the mean scores between the groups. The results do not support hypothesis 2. However, most items had either a significant interaction effect or a marginal interaction effect. Hypothesis 3 was supported by the most of results.

¹ To test hypothesis 1, a questionnaire on the agent, usability, and viewing experience with regard to gender differences was administered. The results indicated that the participants' gender had no significant effect on the evaluation of the agent from experiment 1; social presence towards for males (M = 6.43, SD = 1.45) and females M = 6.58, SD = 1.33; t (34) = -.334 p = .74 (two-tailed); social attraction towards the voice agent for males (M = 6.12, SD = 2.02) and females (M = 6.74, SD = 1.76 t(34) = -.96 p = .34 (two-tailed); and perception of likability about voice agent for males (M = 7.23, SD = 1.21) and females (M = 7.8, SD = 1.36) t (34) = -1.32 p = .19 (two-tailed). The result of the usability test did not show significant differences based on subjects' gender. Consider the following results the usefulness for male (M = 7.55, SD = 1.27) and female M = 8.0, SD = 1.15; t = (34) = -.247 p = .676 (two-tailed); ease of use (M = 8.0, SD = 1.27) and female M = 8.61, SD = 1.07; t = (34) = -1.55 p = .12 (two-tailed); ease of learning (M = 8.63, SD = 1.21) and female M = 8.93, SD = 0.95; t(34) = -.837 p = .40 (two-tailed); satisfaction (M = 7.63, SD = 1.21) and female M = 7.88, SD = 1.46; t = (34) = -.569 p = .57 (two-tailed); buying intention for male (M = 6.94, SD = 1.40) and female M = 7.15, SD = 1.55; t(34) = -.421 p = .676 (two-tailed). Also, consider the following results for viewing experience and content evaluation satisfaction: enjoyable viewing experience (M = 7.00, SD = 1.71) and female M = 7.14, SD = 1.87; t(34) = -.421p = .80 (two-tailed); content evaluation satisfaction (M = 7.81, SD = 1.33) and female M = 7.78, SD = 1.51; t (34) = 0.05 p = 0.95 (two-tailed).

The summary of experiment 1 is that gender has no significant effect on 10 dependent variables.

The evaluation of the agent. The result indicated that the participants' genders had no significant main effect on the evaluation of the two agents (Fs < 1), but the interaction effect between gender and agent was statistically significant on evaluation of the agent and marginally significant with regard to the social presence F(1,32) = 3.16, p < 0.08 multivariate partial eta squared = .09, social attraction F(1,32) = 4.39, p < 0.05 multivariate partial eta squared = .012, and perception of likability F(1,32) = 3.16, p < 0.08 multivariate partial eta squared = .09. According to the results, hypothesis 3 is accepted.

Usability. The same analysis was submitted. The results had no significant main effect ps < .10; however, interaction effects were found for usefulness, F(1,32) = 4.19, p < 0.05 multivariate partial eta squared = .11; satisfaction F(1,32) = 3.19, p < 0.08 multivariate partial eta squared = .012; and buying intention, F(1,32) = 4.23, p < 0.05 multivariate partial eta squared = .12. According to the results, hypothesis 3 is accepted.

Results of Experiment 1

Viewing experience and content evaluation. Similar results were produced: there was no significant main effect. However, a significant interaction effect was found for both viewing experience F(1,32) = 5.84, p < 0.05 multivariate partial eta squared = .04 and content evaluation F(1,32) = 3.10, p < 0.08 multivariate partial eta squared = .08.

5 Conclusion

5.1 Discussion

Is there a difference in how people perceive specialist or generalized voice agents. Are they more comfortable with an agent that is consistent and follows them around platforms or an agent that is specialized for a particular platform? In this study we find no gender differences in how they evaluate one voice agent in one interface. But the answer to the question appears to be that gender has a strong effect on all dimension of how individual perceive a generalized voice agent of a specialized voice agent.

While there was no main effect for the type of voice agent, user's gender significantly affected the social perception of the agent, the agent's usability, and the users perception of the experience and value of the product. Generally, females responded more favorably to a consistent, female agent. They seemed to respond negatively to a new specialized agent, tending to score this new female agent lower. Males, on the other hand, tended to prefer, new specialized female agents.

The interaction effect can be explained by "affect and cognition" – the rational and emotional ways of dealing with social interaction (Forgas 2008). Also, in previous research on computers as social actors individuals show social responses towards computers that are highly influenced by human social categories, especially gender (Nass and Moon 2000; Reeves and Nass 1996; Sundar and Nass 2000; Sundar and Nass 2001).

Affect and Cognition. Considerable research has been done on the interplay between affect and cognition – rational and emotional ways of dealing with objects and

individuals. This distinction (Forgas 2008) has important implications for relational exchanges as encounters between principles and agents (service providers) (Singh and Sirdeshmukh 2000). Affective characteristics of an entity might contain likability and familiar information, whereas cognitive characteristics of a spokesperson might include credibility, expertise, trustworthiness. In general, when social interaction occurs, research suggests that females use a "peripheral route" that uses less of a main message and the affective characteristics, while males who uses "central route" that emphasis on message content.

Categorical Perception. Individuals receive first impressions automatically by relying upon the category to which a social object has been assigned. Categorical perception helps to memory for social information (Brewer 1998). In the cognitive-psychology view, labels have significant effects on how people perceive social objects. Media specialization suggests that people show a categorical perception that relies on the labels assigned to media (Nass et al. 1996). The label or representation of the social category for example specialist label come into a tendency to be biased in their perceptions (Ashforth and Humphrey 1997). Once an object has been categorized, individuals tend to interpret additional cues in line with the categorization, and they may not pay attention to inconsistent information (Hamilton, Sherman & Ruvolo, 1990). For example, when an individual is labeled by definition (e.g., Cindy is a specialist), people perceive him or her based on the central attribute of the social category (e.g., expertise) whether or not that person actually possesses the attribute. That is, the label initiates the central attribute (expertise) assigned to the representation of the category. However, when the research was conducted, the Korean version of Siri had not yet been produced and the smart TV with the voice control system had just emerged. Therefore, participants lacked experience with the machines and the novel attitude could be observed with regard to both voice control systems, Siri and Sori. In sum, this research suggests that there are gender differences as relates to response to two different voice control agents, specialist and generalist (consistent agent,): females gave a better evaluation to the consistent agent, and males preferred the specialist agent.

References

- Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., et al.: A view of cloud computing. Commun. ACM 53(4), 50–58 (2010)
- Ashforth, B.E., Humphrey, R.H.: The ubiquity and potency of labeling organizations. Organ. Sci. **8**(1), 43–58 (1997)
- Berglund, A., Johansson, P.: Using speech and dialogue for interactive TV navigation. Univers. Access Inf. Soc. **3**(3), 224–238 (2004)
- Brewer, M.B.: Category-based vs. person-based perception in intergroup contexts. Eur. Rev. Soc. psychol. **9**(1), 77–106 (1998)
- Brody, L.R., Hall, J.A.: Gender and Emotion. Guilford Press, New York (1993)
- Bullis, C., Horn, C.: Get a little closer: further examination of nonverbal comforting strategies. Commun. Rep. 8(1), 10–17 (1995)

Field, S.: Screenplay: The Foundations of Screenwriting. Bantam Doubleday Dell, New York (1994)

Fiske, S.T., Taylor, S.E.: Social Cognition. McGraw-Hill, New York (1991)

- Forgas, J.P.: Affect and cognition. Perspect. Psychol. Sci. 3(2), 94–101 (2008)
- Hamilton, D.L., Sherman, S.J., Ruvolo, C.M.: Stereotype-based expectancies: effects on information processing and social behavior. J. Soc. Issues 46, 35–60 (1990)
- Hawkins, J.: Computers and girls: rethinking the issues. Sex Roles 13(3-4), 165-180 (1985)
- Kiesler, S., Sproull, L., Eccles, J.S.: Pool halls, chips, and war games: women in the culture of computing. Psychol. Women Q. 9(4), 451–462 (1985)
- Nass, C., Lee, K.M.: Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. J. Exp. Psychol. Appl. 7(3), 171 (2001)
- Nass, C., Reeves, B., Leshner, G.: Technology and roles: a tale of two TVs. J. Commun. 46(2), 121–128 (1996)
- 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)
- Nass, C.I. Steur, J.S., Tauber, E., Reeder, H.: Anthropomorphism, agency, and ethopoeia: computers as social actors. In: Proceedings of the International CHI Conference, Amsterdam, The Netherlands (1993)
- Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. J. Soc. Issues 56 (1), 81–103 (2000)
- Reeves, B., Nass, C.: How people treat computers, television, and new media like real people and places: CSLI Publications and Cambridge University Press (1996)
- Singh, J., Sirdeshmukh, D.: Agency and trust mechanisms in consumer satisfaction and loyalty judgments. J. Acad. Mark. Sci. 28(1), 150–167 (2000)
- Sundar, S.S., Nass, C.: Source orientation in human-computer interaction: programmer, networker, or independent social actor? Commun. Res. 27(6), 683–703 (2000)
- Sundar, S.S., Nass, C.: Conceptualizing sources in online news. J. Commun. 51(1), 52-72 (2001)
- Thomas, F., Johnston, O.: Disney Animation: The Illusion of Life. Abbeville Press, New York (1981)
- Trautschold, M., Ritchie, R., Mazo, G.: Multitasking and Siri. iPhone 4S Made Simple, 179–188 (2011)
- Watson, C.A.: Technology Trends and Predictions: Is a Flying Car in Our Future? (2012)
- Whitley, B.E.: Gender differences in computer-related attitudes and behavior: a meta-analysis. Comput. Hum. Behav. **13**(1), 1–22 (1997)
- Ye, X., Huang, J.: A framework for cloud-based smart home. In: 2011 International Conference on Computer Science and Network Technology (ICCSNT), vol. 2, pp. 894–897. IEEE, December 2011