



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

언론정보학석사학위논문

Expressing the Personalities of the Conversational Agents with Visual and Verbal Feedback and Its Match with Tasks

시각적, 언어적 피드백을 통한 대화형
에이전트의 성격 표현 및 수행 과제에 따른
성격의 선호도

2019년 8월

서울대학교 대학원

언론정보학과

이서영

Expressing the Personalities of the Conversational Agents with Visual and Verbal Feedback and Its Match with Tasks

지도 교수 이 준 환

이 논문을 언론정보학 석사학위논문으로 제출함
2019년 1월

서울대학교 대학원
언론정보학과
이서영

이서영의 언론정보학과 석사학위논문을 인준함
2019년 1월

위 원 장 _____ 홍화정 _____ (인)

부위원장 _____ 서봉원 _____ (인)

위 원 _____ 이준환 _____ (인)

Abstract

Conversational agents with psychological abilities could facilitate natural communication between humans and computers while conversational agents' unnatural expressions and reactions could frustrate users. This research applies the concept of personality to conversational agents to implement natural feedback and reactions.

This study explores how to express conversational agents' personalities. The selected cues were visual feedback and verbal cues. As a between-participants study design, Study 1 measured the perception of five personalities toward different visual feedback and Study 2 measured the perception of five personalities depending on different verbal cues with voices of different genders. Concerning that certain personalities of conversational agents were considered more suitable for certain tasks, Study 3 investigated the user preference and perceived intelligence toward conversational agents with different personalities and tasks.

The study results demonstrate that different motions of visual feedback were highly influential on the perceptions of personalities. Color was not a decisive factor. In addition, except for agreeableness, different verbal cues were perceived as different personalities. For conversational

agents performing service, physical, and office tasks, openness was the most preferred and perceived as intelligent. In case of social tasks, the extravert conversational agents were the most preferred and perceived as intelligent.

Fast and active visual feedback is suitable to design conversational agents with distinct and positive personalities. In addition, perceptions of conversational agent's personalities differed according to the gender of voice. Diverse and expressive cues were suitable for expressing positive personalities. Interactions between conversational agents and humans demonstrated similar patterns of perception as human-human interactions.

Keywords: Personality expression, Conversational agent, Visual feedback, Verbal cues, Task

Student Number: 2017-25742

Table of contents

1. Introduction.....	1
2. Related work.....	6
2.1. Expressing machines' internal states in Human-Computer Interaction.....	6
2.2. Personality expressions of computers and interfaces	8
2.3. Combinations of diverse cues	10
2.4. The agent's personality and task match	11
3. Study 1.....	13
3.1. Overview	13
3.2. Study 1-1	14
3.2.1. Experimental materials	14
3.2.2. Experimental setting	17
3.2.3. Results	18
3.3. Study 1-2	25
3.3.1. Experimental materials	25
3.3.2. Experimental setting	27
3.3.3. Results	28
3.4. Results.....	31

3.5. Discussion	33
3.6. Limitations & Future Studies	37
4. Study 2.....	39
4.1. Overview	39
4.2. Research questions	40
4.3. Method... ..	41
4.3.1. Experimental materials	41
4.3.2. Experimental setting	44
4.4. Results... ..	45
4.4.1. Result 1: Pitch levels and gender of voices	45
4.4.2. Result 2: Emotionality and Gender of voices	47
4.4.3. Result 3: Wordiness and gender of voices.....	53
4.4.4. Result 4: Speed and gender of voices	54
4.4.5. Result 5: Questioning and gender of voices	60
4.5. Overall results	63
4.6. Discussion	65
4.7. Limitations	67
5. Study 3.....	69
5.1. Overview	69
5.2. Method	70

5.2.1. Experimental Materials.....	70
5.2.2. Manipulation check	73
5.2.3. Experimental Setting	74
5.3. Results.	75
5.3.1. Office task.....	75
5.3.2. Social task.....	77
5.3.3. Service task.....	80
5.3.4. Physical task	82
5.4. Discussion	85
6. Conclusions	87
7. Discussion for overall study	91
References	93
Appendix 1. Big Five personality questionnaires	101
Appendix 2. God speed scale questionnaires	102
국문 초록.....	103

Table Index

Table 1 Five visual feedbacks with different levels of motions and colors ..	15
Table 2 Results of one-way ANOVA for five visual feedback	19
Table 3 Results of Tukey HSD <i>post hoc</i> test for visual feedback 1, 2, 3, 4, and 5	21
Table 4 Detailed elements of visual feedback 6, 7, and 8	26
Table 5 Results of ANOVA for visual 6, 7, and 8	29
Table 6 Results of Tukey HSD <i>post hoc</i> test for visual feedback 6, 7, and 8.	30
Table 7 Overall results of perceived personalities for each visual feedback	32
Table 8 Results of two-way ANOVA.....	46
Table 9 Results of Tukey HSD post-hoc test for pitch	47
Table 10 Results of two-way ANOVA for emotionality	48
Table 11 Results of Tukey HSD post-hoc test of emotionality	51
Table 12 Results of two-way ANOVA for wordiness	53
Table 13 Results of Tukey post-hoc test for wordiness	54
Table 14 Results of two-way ANOVA for speech speed	55
Table 15 Results of Tukey post-hoc test for speech speed	58
Table 16 Results of two-way ANOVA for questionnaires	60
Table 17 Results of Tukey post-hoc test for questioning	62

Table 18 Overall results of study 2	63
Table 19 The experimental materials of study 3	71
Table 20 Results of one-way ANOVA for office task	76
Table 21 Results of Tukey HSD post-hoc test for office task	77
Table 22 Results of one-way ANOVA for social task	78
Table 23 Results of Tukey HSD post-hoc test for social task	79
Table 24 Results of One-way ANOVA for service task	80
Table 25 Results of Tukey HSD post-hoc test for service task	81
Table 26 Results of One-way ANOVA for physical task	83
Table 27 Results of Tukey HSD post-hoc test for physical task	84

Figure Index

Figure 1 Visual feedback 1, 2, 3, 4, and 5 used in the study 1-1. 14

Figure 2 Visual feedback 6, 7, and 8 used in the study 1-2. 26

1. Introduction

Future artificial intelligence assistants must be more than just question-and-answer machines. Computers need to express and recognize affect and emotions, because affective computers can help reduce user frustration during interactions and enable comfortable user-computer communication (Picard, 1999).

Computers with unnatural expressions and reactions could deter the human-computer interaction and frustrate users. Therefore, feedback that matches users' perceptions of natural human-computer interaction needs to be designed.

Most studies in the field of affective computing have applied emotion to naturally deliver an agent's internal states. However, this research applies the concept of personality to implement more natural feedback for a conversational agent for the following reasons.

The concept of personality can allow individualized and complex features to be delivered simultaneously. In addition, it is possible to predict users' perceptions by giving the personality to conversational agents. Finally, with personality, agents could implement consistent

patterns of reactions rather than simple and immediate reactions.

Personalities are not innate for computer systems or interfaces; thus, humans must program them. In addition, they cannot be defined in a single sentence; instead, they are combinations of abilities, beliefs, preferences, dispositions, behaviors, and temperamental features with diverse behavioral and emotional attributes (Mairesse & Walker, 2009; Pervin & John, 1999). Therefore, the overall study posits the following research question: *How can conversational agents effectively express personalities?*

Personalities could be expressed with two factors: 1) behavioral features (gesture and movement) and 2) verbal traits (voice and speech style). People with different personalities demonstrate different amplitudes and speeds of gestures and movements. Extraverts demonstrate faster, wider, broader movements than introverts (Neff et al., 2010). In addition, extraverts tend to demonstrate more reactive and faster movements and body gestures than introverts (Brebner, 1985).

Personality and verbal traits are highly interrelated. Extroverts tend to express more emotionality with positive emotions and use fewer formal expressions with agreement and gratitude (Dewaele & Furnham, 1999). Extroverts also demonstrate shorter silences and use more positive

words and informal expressions than introverts. Introverts use more abstract words and formal words than extroverts (Heylighen & Dewaele, 2002; Mairesse et al., 2007).

Gestures are difficult to implement with current conversational agents because most of them are designed in the form of Artificial Intelligence speakers. Instead of using gestures, AI speakers deliver simple visual feedback through the smart display in response to the voice command. Considering the current form of conversational agents, visual feedback was chosen as the personality expression element. As people with different personalities demonstrate different gesture speeds, the speed of lighting could also be perceived as different personalities. Quick lighting is perceived as more active than slow lighting; therefore, the study posits the first Research Question: ***RQ 1. Can different visual feedback be perceived as different personalities?*** Studies 1-1 and 1-2 were conducted to answer this.

Most current AI speakers use a consistent voice with the same speech style. As extroverts and introverts demonstrate different verbal traits, different verbal traits are perceived as different personalities. Therefore, the study posits the second Research Question: ***RQ 2. Can different verbal cues be perceived as different personalities?*** Study 2

was conducted to answer this research question.

Putting personalities into machines or computers need not have a time or space cost because a personality could be applied to almost anything that presents itself to users (Nass et al., 1995). In addition, personalities themselves are a reduced basic set of categories. The goal of this research is to find the minimum set of necessary cues to present a computer-based personality, focusing on the use of specific cues to express the personalities of conversational agents.

Matching a conversational agent's personality with its purpose enhances its functionality and job performance. This is supported by the understanding that certain personalities are better suited for particular tasks (Barrick & Mount, 1991; McCrae & Costa, 1997), and the personality expressed by a social agent can influence users' perception of how well the agent has performed a particular task (Liao et al., 2018). Applying certain personalities to certain agent's tasks is required as it provides consistency and ease of understanding in Human-Computer interaction (Kwak, 2014). Therefore, the third research question is: ***RQ 3. Which personality do users prefer for which tasks performed by a conversational agent?*** Study 3 was conducted to answer this research question.

Unlike previous studies, the current study applies a wider range of personalities rather than focus on expressing only two contrasting personalities such as extraversion and introversion. In addition, the personality is expressed with diverse factors including visual feedback with different colors and motions and five verbal cues rather than focusing only on a single element.

Studies 1 and 2 aimed to provide the optimal verbal cues and visual feedback for expressing the personalities of the conversational agents and provide design implications. With reference to the results of studies 1, 2, and 3 will examine the user preferences and perceived intelligence toward the conversational agent performing different tasks with different personalities.

2. Related work

2.1. Expressing machines' internal states in Human-Computer Interaction

People treat computers as the social actors (Reeves & Nass, 1996) and apply the rules of human interactions. For instance, people prefer flattering computers that praise them to cynical computers that criticize them. Computers that perform like humans are required and people treat computers as social actors (Reeves & Nass, 1996). Developing computer agents that are capable of performing like humans requires that computer systems have social competencies designed into them (Bickmore & Cassell, 2005).

Considering the uncertainty of systems, it is difficult to set absolute standards for expressing the machine's psychological and emotional states. According to previous studies, in the field of Human-Computer Interaction, machines' internal states can be delivered and expressed through its diverse elements.

Verbal cues and texts are the most direct ways to deliver machines' and computers' internal states. Voice-related factors such as

volume, frequency range, and speech rate are associated with the internal states of machines and computers (Aly & Tapus, 2013; Lee et al., 2006).

Non-verbal behaviors such as facial expressions, interpersonal closeness, gestures (Aly & Tapus, 2013), and gaze (Bickmore & Cassell, 2005; Oliveira et al., 2018) of embodied conversational agents were adjusted to deliver its social competencies in conversations. In particular, the gaze of the conversational agents was used to deliver the backchannel feedback in turn-taking situations (Torres et al., 1997).

Diverse colors and lights were also used to express the social robots' emotions (Löffler et al., 2018; Song & Yamada, 2017; Terada et al., 2012) and a simple blinking light was used as a subtle expression to smooth human-robot speech interactions (Funakoshi et al., 2008).

This study argues that, for the long-term of human-computer interaction, the interface's personality plays a more crucial role than simple emotional reactions do. The current approach for personality expression was adapted from previous affective computing (Picard, 1998; 1999) and expressive output modalities studies (Song & Yamada, 2017; Terada et al., 2012). For instance, similar to prior studies, this study adopted cues to express conversational agents' internal states such as personality. Affective computers can both perform better when assisting

humans and enhance their decision-making skills.

2.2. Personality expressions of computers and interfaces

Personality is defined as consistent patterns of feeling, thinking, and behavior (Pervin & John, 1999). Among diverse expressive factors, personality was chosen to express the internal states of conversational agents in this study for the following reasons: First, personality can easily express the subtle and complex dimensions in individual differences. Second, personality can provide a wide range of predictions concerning how people may respond to various personalities (Nass & Lee, 2000). In other words, people unconsciously use personality as a tool to assess the social partners (Dryer, 1999). Also, much literature support that individuals assess computers according to personality (Isbister & Nass, 2000; Nass, 2000; Nass et al., 1995). Lastly, personality is essential (Breazeal, 2004; Lee et al., 2006) in that it can help agents build natural relationships with humans.

Personality can be expressed and perceived in diverse ways because diverse viewpoints exist regarding personality. Personality could

be seen as only measurable consistencies in behavior, or personality could be seen as perceived consistencies regardless of its actual measurable consistencies. Among diverse perspectives, the current study argues that personality is expressed and perceived through the interactions between the expectations of observers and the behaviors of the observed (Dryer, 1999).

Most prior studies on machine's expressions of personality particularly have focused specifically on expressing two contrasting personalities such as extraversion/ introversion or dominance/ submissiveness. Comparatively few studies of personality expression have applied diverse types of personalities such as openness and conscientiousness.

Among diverse categorizations of personalities, the Big Five personality has earned high reliability and validity in psychology and communication studies (Hancock & Dunham, 2001; Reeves & Nass, 1996; Wang et al., 2015). The revised version of the Big Five NEO (Neuroticism-Extraversion-Openness) personality inventories became more comprehensive versions of personality models by including agreeableness and conscientiousness (Costa & McCrae, 2017). Since the current study argues that more sophisticated and complex expressions of

conversational agents are needed, the Big Five personality was adopted rather than focusing on two personalities.

The detailed descriptions for each Big Five personality are as follows. Agreeableness is associated with being friendly, courteous, flexible, trusting, good-natured, cooperative, forgiving, soft-hearted, and tolerant. Extraversion is associated with being sociable, gregarious, assertive, talkative, actively ambitious, and initiative. Conscientiousness is known as conformity or dependability, with careful, thorough, responsible, organized, planning, and hardworking traits. Openness is related to intellect or intelligence, which includes traits such as imaginative, cultured, curious, original, broad-minded, intelligent, and artistically sensitive. Neuroticism is also known as emotionally unstable and is related to the degree of negative emotion, such as being moody and tense (Costa & McCrae, 2017; Rammstedt & John, 2007).

2.3. Combinations of diverse cues

Two representative environmental factors in affective computing are 1) visual and behavioral elements (including color, light, video, and animations), and 2) audio stimuli (including noise, sound, and music).

The current study uses both visual and verbal feedback.

Previous studies effectively examined personalities using diverse interfaces and contexts. Behavioral and linguistic factors and their correlations with personality were observed in virtual environments (Yee et al., 2011). A social robot using different postures, gestures, and eye gazes, depending on its personality and that of its user was also studied (Aly & Tapus, 2013; Andrist et al., 2015). Linguistic cues and their correlations with personality (Nass & Lee, 2000) were also studied. LED lights (Sokolova & Fernández-Caballero, 2015) were used for the internal expressions of interfaces.

Based on previous studies, the current study argues that the personality of interfaces must be expressed using combinations of diverse cues. Therefore, both visual feedback and paralinguistic and verbal cues were adopted to express the interface's personality.

2.4. The agent's personality and task match

Tasks that personal Artificial Intelligent assistants perform are highly related with their personalities. In other words, people expect different personalities depending on the agent's functional roles and tasks

(Joosse et al., 2013; Lee et al., 2018; Tay et al., 2014). In previous studies about chatbot, people expect chatbot personality to match with its tasks. People expected a news chatbot to be professional, while a shopping chatbot to be humorous and casual (Jain et al., 2018).

Previous studies have applied various standards to categorize robot tasks. Depending on how they perform jobs, tasks can be classified as objective or subjective. Depending on the actual work that robots perform (Bartneck et al., 2009).

This study applied the categorization of tasks referring to the previous studies, which are based on people's expectations of robot tasks (Lee et al., 2018). Office tasks include copying, typing, and printing; Physical tasks include household chores, physical labor, and security; Service tasks are guiding, serving in restaurants and shops, and sales. Social tasks include making conversation, entertaining, and telling jokes.

This study argues that users have certain expectations for psychological expressions of their agents that should match the tasks they perform. Therefore, this study aims to determine preferred personalities depending on tasks that agents perform by applying previous studies' task categorizations.

3. Study 1

3.1. Overview

Motion and color were chosen as the elements of personality expressions. As behavioral traits such as gestures and movements play important roles in personality perception (Brebner, 1985; Neff et al., 2010) the speed of motion was chosen as the element of visual feedback. In addition, color was chosen to express personalities of conversational agents because light and color are essential factors in affective communication between users and computers, particularly in computer interfaces (Adams et al., 2015; Löffler et al., 2018; Song & Yamada, 2017; Terada et al., 2012).

Visual appearances play a key role in the perceptions of personalities (Dryer, 1999). Considering that personalities can be applied everywhere with diverse forms, colors and motions were chosen (Reeves & Nass, 1996). Color and motion are readily applicable, quick, and low-cost, and the most common output modalities used in diverse interfaces such as appearance-constrained social robots (Löffler et al., 2018).

Study 1 will examine the personality perception of five visual feedback with different colors and motions. To answer *RQ1. Can*

different visual feedback with different colors and motions express different personalities? Studies 1-1 and 1-2 were conducted.

3.2. Study 1-1

3.2.1. Experimental materials

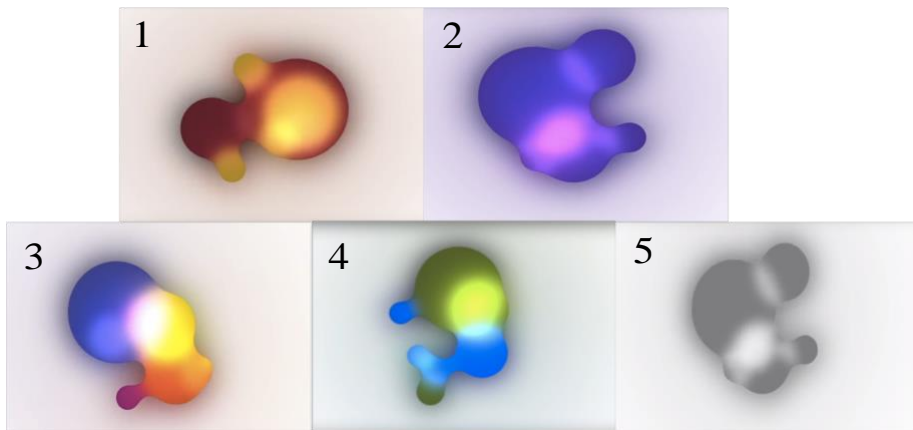


Figure 1 Visual feedback 1, 2, 3, 4, and 5 used in the study 1-1

Table 1 Five visual feedbacks with different levels of motions and colors

Visual feedback	Motions (slow, moderate, fast)	Colors (yellow, purple, red, blue, green, grey)
1	Moderate	Red + Yellow
2	Slow	Purple
3	Fast	Red + Blue + Yellow
4	Moderate	Green+ Blue
5	Slow	Grey

The main object of visual feedback is abstract moving circle. The design of visual feedback was motivated by diverse traits of Big Five personalities. Concerning that the Big Five personality traits are measured with bipolar adjectives (Costa & McCrae, 2017), color were chosen in the color wheel diagnostically, vertically, and horizontally. Expressivity was reflected as different number of colors in the visual feedback. This is because Big five personalities demonstrate different levels of expressivity (Costa & McCrae, 2017).

When using multiple colors, colors were only partially blended and overlapped, not entirely mixed. The partial overlapping of colors was motivated by the concept of continuity since each personality is defined as continuous traits between bipolar adjectives.

The motion was chosen as the design element of visual feedback because five personalities have distinctly different levels of expressivity, proactivity, and energy (McCrae & Costa, 1997). The speed of motions was adjusted using Adobe Premiere Effect program, the speed of movements was adjusted between fast (0.35), moderate (0.25), and low (0.15) with the program. The size of motions was also adjusted in parallel with the speed of motions. Visual feedback with fast motion demonstrates wider movements. In total, five visual feedback were designed with different colors and motions.

The main object was the abstract circle format and its colors and motions were adjusted. When using multiple colors, they were only partially mixed and blended, but distinct color differences were still demonstrated. The background to all visual feedback was a white blank space.

3.2.2. Experimental setting

In total, 30 participants (14 females and 16 males) aged 20–35 years were recruited for the experiment ($M = 28.8$, $SD = 3.4$). There were six participants in each of the five between-participants conditions, which were visual feedback with different colors and motions.

Participants entered the room and received oral instructions about the study. They watched one visual feedback video that combined different colors and motions and answered survey questions about the short version of the Big Five personality questionnaires (Rammstedt & John, 2007) to see which visual feedback factors were best suited for the five personalities. In total, 10 questions were asked of participants in the survey. The detailed questionnaires of the Big Five personalities are attached (Appendix 1).

A telepresence robot was used during the study. A telepresence robot was chosen as the conversational agent for the study because it is suitable for demonstrating both visual feedback and verbal cues. The visual feedback was demonstrated to

participants through the telepresence robot's screen. The height of the telepresence robot was fixed throughout the study.

The results of the experiment were analyzed using one-way ANOVA by comparing the mean values of the personality perceptions depending on the different visual feedback. More detailed differences of between-group personality perceptions were observed using the Tukey HSD *post hoc* test. The next section details the results of the experiment.

3.2.3. Results

One-way ANOVA was conducted for the color and motion variables of each personality to compare the mean values of visual feedback depending on the different personalities. The results showed that different visual feedback was perceived as different personalities, which answered the research question about whether different visual feedback could express different personalities.

The following are detailed statistical results of one-way ANOVA (Table 2). For agreeableness, visual feedback 4 ($F(4, 30) = 14.37, p < 0.001$) demonstrated the highest mean value ($M =$

6.16). For conscientiousness, visual feedback 2 ($F(4, 30) = 6.12, p < 0.01$) demonstrated the highest mean value ($M = 5.83$). For extraversion, visual feedback 3 demonstrated the highest mean value (Mean = 5.9) ($F(4, 30) = 11.01, p < 0.001$). For neuroticism, visual feedback 2 demonstrated the highest value ($M = 6.167$) ($F(4, 30) = 23.15, p < 0.0001$). For openness, visual feedback 4 demonstrated the highest mean value ($M = 3.66$) ($F(4, 30) = 3.24, p < 0.05$). Table 2 explains the statistical results of one-way ANOVA in more detail.

Table 2 Results of one-way ANOVA for five visual feedback

Personality Traits	Visual Feedback	Mean	SD	df	F value	P value
Agreeableness	1	5.833	0.75	4	14.37	.000***
	2	2.667	1.21			
	3	3.667	0.81			
	4	6.167	0.75			
	5	3.333	1.37			
Conscientiousness	1	3.5	1.05	4	6.126	.001**
	2	5.833	0.75			
	3	2.33	1.37			
	4	3.66	1.03			
	5	3.33	1.86			

Extraversion	1	3.5	0.55	4	11.01	.000***
	2	2.83	0.41			
	3	5.9	0.89			
	4	4.33	1.21			
	5	3.33	1.21			
Neuroticism	1	2.00	0.89	4	23.15	.000***
	2	6.167	0.75			
	3	3.667	1.03			
	4	2.833	0.75			
	5	5.833	1.16			
Openness	1	3.16	1.21	4	3.242	0.03*
	2	2.66	0.82			
	3	3.16	0.75			
	4	3.66	0.82			
	5	2.16	0.75			

Tukey HSD *post hoc* test was conducted to see the detailed differences in the mean values of each personality for the five instances of visual feedback (Table 3). For agreeableness, visual feedback 1 demonstrated the significant differences from both visual feedback 2 and 3. For conscientiousness, visual feedback 2 (M= 5.833) was more perceived as conscientious than visual feedback 3 (M = 2.33). Visual feedback 4 (M = 3.67) was more perceived as openness than visual feedback 5 (M = 2.17). Visual

feedback 3 (M = 5.90) was more perceived as extraverted than visual feedback 2 (M = 2.83) and 5 (M = 3.33). Visual feedback 2 (M = 6.16) was more perceived as neurotic than visual feedback 4 (M = 2.83) and 1 (M = 2.00). Visual feedback 4 (M = 3.66) was more perceived as openness than visual feedback 1 (M = 3.16) and 4 (M = 3.66). Table 3 contains more detailed results of the Tukey *post hoc* test.

Table 3 Results of Tukey HSD *post hoc* test for visual feedback 1, 2, 3, 4, and 5

Personality traits	Pair-wise comparisons	Difference	95% confidence level		P- value
			lower	upper	
Agreeableness	Visual feedback 2- Visual feedback 1	-3.16	-4.88	-1.44	0.00***
	Visual feedback 3- Visual feedback 1	-2.16	-3.88	-0.44	0.00**
	Visual feedback 4- Visual feedback 1	0.33	-1.38	2.05	0.97
	Visual feedback 5- Visual feedback 1	-2.50	-4.21	-0.78	0.00**
	Visual feedback 3- Visual feedback 2	1.00	-0.71	2.71	0.44
	Visual feedback 4-	3.50	1.78	5.21	0.00***

	Visual feedback 2				
	Visual feedback 5- Visual feedback 2	0.66	-1.05	2.38	0.78
	Visual feedback 4- Visual feedback 3	2.50	0.78	4.21	0.00**
	Visual feedback 5- Visual feedback3	-0.33	-2.05	1.38	0.97
	Visual feedback5- Visual feedback4	-2.83	-4.55	-1.11	0.00***
Conscientio usness	Visual feedback2- Visual feedback1	2.33	0.18	4.48	0.02*
	Visual feedback3- Visual feedback1	-1.16	-3.32	0.98	0.51
	Visual feedback4- Visual feedback1	0.16	-1.98	2.32	0.99
	Visual feedback5- Visual feedback1	-0.16	-2.32	1.98	0.99
	Visual feedback3- Visual feedback2	-3.50	-5.65	-1.34	0.00**
	Visual feedback4- Visual feedback2	-2.16	-4.32	-0.01	0.04*
	Visual feedback5- Visual feedback2	-2.50	-4.65	-0.34	0.01*

	Visual feedback4- Visual feedback3	1.33	-0.82	3.48	0.38
	Visual feedback5- Visual feedback3	1.00	-1.15	3.15	0.65
	Visual feedback5- Visual feedback4	-0.33	-2.48	1.82	0.99
Extraversion	Visual feedback2- Visual feedback1	-0.66	-2.22	0.88	0.71
	Visual feedback3- Visual feedback1	2.50	0.94	4.05	0.00**
	Visual feedback4- Visual feedback1	0.83	-0.72	2.38	0.52
	Visual feedback5- Visual feedback1	-0.16	-1.72	1.38	0.99
	Visual feedback3- Visual feedback2	3.16	1.61	4.72	0.00***
	Visual feedback4- Visual feedback2	1.50	-0.05	3.05	0.06*
	Visual feedback5- Visual feedback2	0.50	-1.05	2.05	0.87
	Visual feedback4- Visual feedback3	-1.66	-3.22	-0.11	0.03*
	Visual feedback5- Visual feedback3	-2.66	-4.22	-1.11	0.00***
	Visual feedback5- Visual feedback4	-1.00	-2.55	0.55	0.34
Neuroticism	Visual feedback2- Visual feedback1	4.16	2.58	5.751	0.00***

	Visual feedback3- Visual feedback1	1.66	0.08	3.25	0.03*
	Visual feedback4- Visual feedback1	0.83	-0.75	2.41	0.54
	Visual feedback5- Visual feedback1	3.83	2.24	5.41	0.00***
	Visual feedback3- Visual feedback2	-2.50	-4.08	-0.91	0.00**
	Visual feedback4- Visual feedback2	-3.33	-4.91	-1.74	0.00***
	Visual feedback5- Visual feedback2	-0.33	-1.91	1.25	0.97
	Visual feedback4- Visual feedback3	-0.83	-2.41	0.75	0.54
	Visual feedback5- Visual feedback3	2.16	0.58	3.75	0.00**
	Visual feedback5- Visual feedback4	3.00	1.41	4.58	0.00***
Openness	Visual feedback2- Visual feedback1	-1.00	-2.50	0.50	0.31
	Visual feedback3- Visual feedback1	-0.50	-2.00	1.00	0.86
	Visual feedback4- Visual feedback1	0.00	-1.50	1.50	1.00
	Visual feedback5- Visual feedback1	-1.50	-3.00	0.00	0.05
	Visual feedback3- Visual feedback2	0.50	-1.00	2.00	0.86
	Visual feedback4- Visual feedback2	1.00	-0.50	2.50	0.31

Visual feedback5- Visual feedback2	-0.50	-2.00	1.00	0.86
Visual feedback4- Visual feedback3	0.50	-1.00	2.00	0.86
Visual feedback5- Visual feedback3	-1.00	-2.50	0.50	0.31
Visual feedback5- Visual feedback4	-1.50	-3.00	0.004	0.05*

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

3.3. Study 1-2

3.3.1. Experimental materials

Study 1-2 was added to see more obvious patterns of the perceptions of personalities depending on different colors to supplement the results of study 1-1. Motivated by different levels of proactivity and energy of five personalities (McCrae & Costa, 1997), the speed of motions was adjusted in three levels: fast (0.35), moderate (0.25), and slow (0.15). Size of motions was differentiated depending on the speed of motions.

Motivated by different expressivity of five personalities, the number of colors was chosen. Meanwhile, to effectively see the impacts of colors, experimental materials were designed with

different colors from study 1-1 by applying red, blue, and a combination of dark blue, yellow, and orange. In case of color combinations, the colors were not fully mixed or blended, but they all demonstrated partial overlapping, which applied the same design methods with the experimental materials of study 1-1. Detailed information of colors and motions for visual feedback 6, 7, and 8 are demonstrated (Table 4).

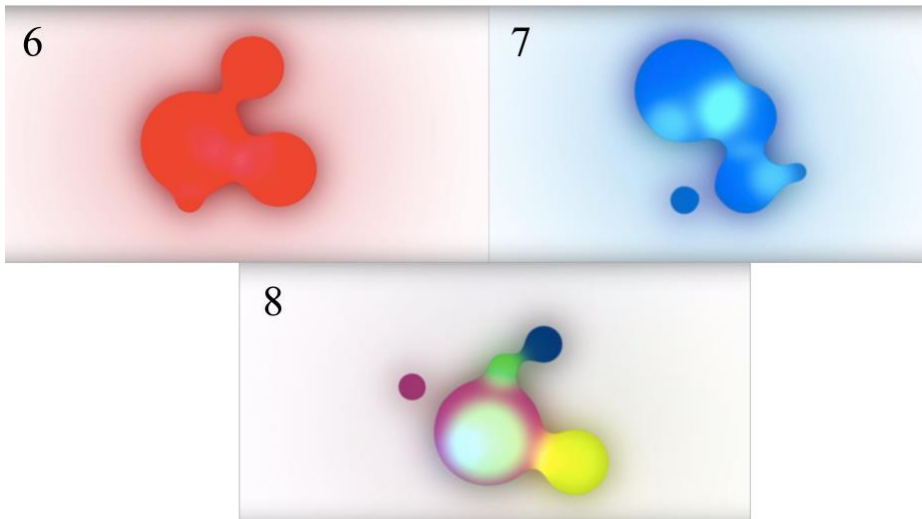


Figure 2 Visual feedback 6, 7, and 8 used in the study 1-2.

Table 4 Detailed elements of visual feedback 6, 7, and 8

Visual	Motion	Colors
--------	--------	--------

feedback	(slow, moderate, fast)	(red, blue, yellow, dark blue, purple)
6	Moderate	Red
7	Slow	Blue
8	Fast	Dark Blue + Yellow + Purple

3.3.2. Experimental setting

Five participants were recruited for each of the three visual feedback instances. Total of 15 participants (10 female and 5 male) aged 20–30 years old were recruited for the experiment ($M = 25.5$, $SD = 3.2$) as between-subjects. Each group of participants was exposed to each video of different visual feedback. After watching the video, they answered ten questions to measure the Big Five personalities.

Participants entered the room and received the oral instructions about the study. They watched one visual feedback video that combined different colors and motions and answered survey questions of the short version of the Big Five personality questionnaires (Rammstedt & John, 2007) to see which visual feedback was the most suitable for the five personalities. In total,

participants were asked 10 questions in the format of survey. The detailed questionnaires of the Big Five personalities are attached (Appendix 1).

The results of the experiment were analyzed using one-way ANOVA by comparing the mean values of the personality perceptions depending on the different visual feedback. After conducting the one-way ANOVA, a Tukey HSD *post hoc* test was conducted to compare more detailed differences between the personality perceptions between visual feedback 6, 7, and 8. The next section demonstrates the detailed results of the experiment.

3.3.3. Results

Table 5 demonstrates the detailed results of one-way ANOVA. Visual feedback 6, 7, and 8 only showed significant differences in the perception of extraversion ($F(2, 15) = 17.90, p < 0.001$) and openness ($F(2, 15) = 28.97, p < 0.001$). Perceptions of other personality types did not show significant differences depending on visual feedback.

Table 5 Results of ANOVA for visual 6, 7, and 8

Personality traits	Visual feedbacks	Mean	SD	df	F value	P-value
Agreeableness	6	4.2	1.25	2	0.23	.79
	7	3.9	1.51			
	8	3.7	0.44			
Conscientiousness	6	3.4	1.19	2	0.97	.40
	7	4.2	0.83			
	8	3.4	1.08			
Extraversion	6	2.1	0.54	2	17.90	.00***
	7	3.6	1.94			
	8	6.6	0.54			
Neuroticism	6	3.2	0.83	2	0.85	.44
	7	2.7	0.97			
	8	2.5	0.79			
Openness	6	2.0	1.00	2	28.97	.00***
	7	1.8	0.83			
	8	6.4	1.34			

A Tukey HSD *post hoc* test was conducted to compare the mean values of personalities between three visual feedback instances. For extraversion, visual feedback 8 (M = 6.6) demonstrated significant differences from visual feedback 6 (M = 2.1). In addition, visual feedback 8 and 7 (M = 3.6) showed

differences. For openness, visual feedback 8 demonstrated significant differences from both visual feedback 7 (M = 1.8) and 6 (M = 2.0). Other personalities did not show differences depending on different visual feedback. Table 6 contains more detailed *post hoc* results.

Table 6 Results of Tukey HSD *post hoc* test for visual feedback 6, 7, and 8.

Personality traits	Family-wise comparisons	Difference	95% confidence level		p value
			lower	upper	
Agreeableness	Visual feedback 7 - Visual feedback 6	-0.30	-2.26	1.66	.91
	Visual feedback 8 - Visual feedback 6	-0.50	-2.46	1.46	.78
	Visual feedback 8 - Visual feedback 7	-0.20	-2.16	1.76	.96
Conscientiousness	Visual feedback 7 - Visual feedback 6	0.80	-0.97	2.57	.47
	Visual feedback 8 - Visual feedback 6	0.00	-1.77	1.77	1.00
	Visual feedback 8 - Visual feedback 7	-0.80	-2.57	0.97	.47
Extraversion	Visual feedback 7 - Visual feedback 6	1.50	-0.54	3.54	.16

rsion	Visual feedback 8 - Visual feedback 6	4.50	2.45	6.54	.00***
	Visual feedback 8 – Visual feedback 7	3.00	0.95	5.04	.00**
Neuroti cism	Visual feedback 7 - Visual feedback 6	-0.50	-1.96	0.96	.64
	Visual feedback 8 - Visual feedback 6	-0.70	-2.16	0.76	.43
	Visual feedback 8 - Visual feedback 7	-0.20	-1.66	1.26	.93
Openne ss	Visual feedback 7 - Visual feedback 6	-0.20	-2.02	1.62	.95
	Visual feedback 8 - Visual feedback 6	4.40	2.57	6.22	.00***
	Visual feedback 8 - Visual feedback 7	4.60	2.77	6.42	.00***

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

3.4. Results

Considering all results (Tables 2–6), different motions of visual feedback were highly influential on the perceptions of personalities. Fast, moderate, and slow motions could be used to express different personalities, which answers the research question about whether visual feedback with different motions could express different personalities.

Regardless of colors, visual feedback with fast and moderate speed of motions were perceived as agreeable and kind. Visual feedback with slow motions was perceived as deliberate and careful. Visual feedback with fast and moderate motions was considered as having a creative and imaginative personality. Visual feedback with fast motions was considered active and sociable. Visual feedback with slow motions was considered depressed and anxious. Table 7 details the overall results for the visual feedback that are appropriate to express personalities.

Table 7 Overall results of perceived personalities for each visual feedback

Visual feedback	Speed of motions (slow, moderate, fast)	Personality traits
1	Moderate	Openness
2	Slow	Conscientiousness, Neuroticism
3	Fast	Extraversion, Agreeableness
4	Moderate	Agreeableness, Openness

5	Slow	Neuroticism
8	Fast	Extraversion, Openness

3.5. Discussion

The traits of extraversion are related to ambition and sociability. Agreeableness can be interpreted as likability, cooperation, social conformity, and love (Barrick & Mount, 1991) and these traits are more highly related with positive personalities than the rest of the Big Five personalities. The study results suggest that fast motions are appropriate for expressing positive personalities such as agreeableness and extraversion.

The results support the previous finding that users' perceptions of extraversion increase as the motion level increases (Hyde et al., 2013). Concerning that previous studies about motions were conducted with virtual characters' behavioral movements, this study provides wider design implications for more diverse formats of appearance constraint agents with active personalities.

The study argues that using slow motions are suitable for

personalities that are usually perceived as negative concerning the results of personality perceptions associated with slow motions (Table 7). The results present that slow motions were perceived as neurotic and conscientious. According to the Big Five personality trait study (McCrae & Costa, 1997), traits associated with Neuroticism are highly related to emotional instability, anxiety, and insecurity. In addition, traits associated with Conscientiousness are related to thoroughness and planning, which could be negatively perceived in relationships.

Fast and active visual feedback is required to design conversational agents with distinct and positive personalities such as extraversion, agreeableness, and openness. This is similar to human–human conversations because actively reactive conversational partners are more perceived as having active and positive personalities than passive and sullen conversational partners.

The study results demonstrate that color is not a valid factor to express different personality traits. Colors do not show any consistent or significant patterns depending on different personality traits. Even though the two studies used experimental materials with

diverse color hues such as red, purple, green, and green with blue, consistent patterns of personality perceptions were only demonstrated depending on the speed of motions, regardless of color.

This study's results bring up the issue of color subjectivity and objectivity. The subjectivity and objectivity of colors is a highly controversial issue among color scientists. Based on the argument that color perceptions could be organized with objective standards and systems, the relationships between colors and personalities have been widely observed and studied, particularly in human psychology. For instance, personality testing systems such as the Color Pyramid Test (Schaie, 1963) exist to evaluate personalities depending on different colors. In addition, according to the theory of color (Warner, 1966), red, yellow, and orange are related to exciting and enlivening features, blue and purple are related to anxiety and yearning features, yellow is related to anger, and black is related to depression. Bold colors are more suitable for expressing dominant personalities than submissive personalities (Dryer, 1999).

The current study argues that colors are not suitable for

expressing diverse personalities because personal preference is decisive in the perception of colors. The results of this study that colors are not influential factors could be supported by diverse color perception studies. Color eliminativism, which is the notion that physical objects are not colored at all and colors are perceived psychologically rather than objectively, supports the current study's results (Byrne & Hilbert, 2003). In addition, the color researcher Jean-Philippe Lenclos's study of geographical color demonstrated that colors could be explained and defined based on geographical, cultural, and geographical conditions (Lenclos, 2004), which is highly related with the ecological perspective on color (Byrne & Hilbert, 2003). The correlations between specific colors and psychological factors as neither objective nor complicated (Brave & Nass, 2003); therefore, more sophisticated color schemes that reflect other design elements should be designed for future studies.

The results could cautiously provide the slight possibility that the number of colors could influence personality perceptions rather than their hue. In particular, both visual feedback instances 3 and 8 demonstrated significantly high mean values compared with other visual feedback instances in extraversion (Tables 2 and 5).

Considering that only visual feedback instances 3 and 8 used three colors (Tables 1 and 4), while the others used only one or two colors, we could suggest that using diverse colors rather than a single color could be suitable for expressing extraversion. However, since this study cannot ignore the impacts of hues, this study result does not fully support this argument.

3.6. Limitations & Future Studies

More carefully designed color schemes combined with diverse design elements are needed. This study lacked a logical argument for choosing colors. Therefore, if future studies are conducted in the same experimental setting, the colors need to be constrained because the perceptions of colors are highly related to cultural standards and individual preferences. If the study is conducted based on the argument that colors could be objects' objective physical properties, colors need to be chosen based on more objective standards of color systems such as the Natural Color

System¹ or Munsell Color System (Munsell, 2018).

In addition, future studies need to demonstrate more diverse and detailed levels of speed. The current study focused solely on three speed levels: fast, moderate, and slow, but a wider range should be considered to express diverse ranges of personalities. The study result showed that slow motions could express both conscientious and neurotic personalities, which could be problematic because these two personalities have distinct differences. This might be caused by the limitation to three levels; therefore, future studies need more detailed motion speed levels.

¹ <https://ncscolour.com/>

4. Study 2

4.1. Overview

Personality and verbal traits such as voice pitch, speed of speech, wordiness, questioning, and voice emotionality are highly related. Study 2 was motivated by previous findings in human communication claiming that people with different personalities such as extraversion and introversion usually demonstrate different verbal traits.

Extraverts and introverts use different verbal cues. Regarding speech speed, extroverts speak faster, and with louder voices, higher pitch, and less monotonous voices than introverts do (Nass & Lee, 2000). Extroverts tend to express more emotionality with positive emotions and use fewer formal expressions with agreement and gratitude (Dewaele & Furnham, 1999). Extroverts also demonstrate shorter silences and use more positive words and informal expressions than introverts do. Introverts use more abstract words and formal words than extroverts do (Heylighen & Dewaele, 2002). Introverts also use more negations and negative emotion words than extraverts do. Motivated by these findings, the study applied verbal cues to express

conversational agents' personalities.

A total of five verbal cues were chosen for the following reasons. Pitch and speed of speech were chosen because these two elements are the key factors of voices communication (Apple et al., 1979) to deliver individual features. Emotionality of voice, which was expressed as the emotional reactions and feedbacks of the conversational agents in this study, was selected, because the agent generating the desired emotional state and personality is important in users' perceptions of the agent (Picard, 1999). The questioning element was chosen based on a previous interpersonal communication study in which the usages of questionnaires differentiate the perception of voices (Apple et al., 1979). The wordiness was chosen by reference to the natural language generator named PERSONAGE (Aly & Tapus, 2013) that uses verbosity as the element to express personality dimensions through language. With five verbal cues, gender of voice was also chosen as one of the verbal cues because it is one of important elements of personal attributes.

4.2. Research questions

The research question of the study 2 is ***RQ2. "Can different verbal***

cues be perceived as different personalities?”

More detailed research questions for each verbal cue are as follows.

RQ 2-1. Would different levels of pitch and gender of voices be perceived as different personalities?

RQ 2-2. Would different levels of emotional verbal feedbacks and gender of voices be perceived as different personalities?

RQ 2-3. Would different wordy speech style and gender of voices be perceived as different personalities?

RQ 2-4. Would different speed speech rate and gender of voices be perceived as different personalities?

RQ 2-5. Would different usage of questionnaires be perceived as different personalities?

4.3. Method

4.3.1. Experimental materials

Table 3 Experimental materials of verbal cues used in the study 2.

Script	Verbal Cues
1	3 Pitch (high/ moderate/ low) x Gender (female/male)
2	2 Wordiness (high/ low) x Gender (female/male)
3	3 Speed (high/ moderate/ low) x Gender (female/male)
4	2 Emotionality (high/ low) x Gender (female/male)
5	2 Questioning styles (high/low) x Gender (female/male)

Each script was written based on small talk from chatbot Mitsuku² and current AI speakers such as Google Home³ and Amazon Alexa⁴. The contents that commonly appeared in current AI speakers and chatbots as default setting were applied to see the impacts of verbal cues, avoiding heavy influence from contents. Two Oddcast TTS programs were used for the manipulations of voices. Different voices were designed using the PC internal recorder with different scripts.

Manipulation of each verbal cue occurred as follows. The pitch level was adjusted using the Audacity program⁵ and Oddcast TTS⁶.

² <https://www.pandorabots.com/mitsuku/>

³ https://store.google.com/gb/product/google_home

⁴ <https://www.amazon.com/Amazon-Echo-And-Alexa>

⁵ <https://www.audacityteam.org>

Using the Amazon Polly female and male voices as the default voice, the pitch level of voices was adjusted. The effects of voice pitch on the evaluation of a social robot was examined (Niculescu et al., 2011). Exuberant voice and calm voice were manipulated with different pitch levels. In case of male voice, high pitch was 125.5 Hz, moderate pitch was 110 Hz, and low pitch was 98 Hz. In case of female voice, high pitch was 226 Hz, moderate pitch was 213 Hz, and low pitch was 200 Hz. The pitch level was adjusted based on the previous studies about pitch and personality (Apple et al., 1979; Niculescu et al., 2011). Script 1 asked a social robot to introduce itself.

The wordiness was manipulated with two level: high and low. The wordiness was adjusted with the length of the conversational agent's answers. High wordiness was manipulated as more than 20 words and low wordiness condition was manipulated as less than 20 words. Script 2 asked for a name of the robot.

The speed was manipulated with three levels using the Audacity program. For female voices, high speed was 18.5 times and low speed was -11 times. For male voices, high speed was 20 times and low speed

⁶ <http://www.oddcast.com/demos/tts/emotion.html#>

was -11 times. Voices with moderate speed were not adjusted by the program. Script 3 asked the weather conditions to the robot.

The emotionality of the voice was manipulated with two levels. For voices with emotionality, the Oddcast program with emotion functions was applied. Emotional expressions such as “Aha” and “Wow” were included. The Oddcast program without emotion was applied for voices without emotionality. Script 4 was to give directions. The participant asks the agent for directions from home to Gangnam station.

The questionnaire condition was manipulated with two levels: high and low. High level of questionnaire was expressed in the format of interrogative sentence. Low level of questionnaire was expressed in the format of declarative sentence.

4.3.2. Experimental setting

For study 2, a total of 60 participants were recruited for the experiment. Subjects were divided into five groups; each of the 12 participants in each group heard different voices with different conditions of verbal cues with the same script. After that, participants answered the short version of the Big 5 personality survey questions

(Appendix 1) for each verbal cue using a seven-point Likert scale.

The telepresence robot was used. The height was fixed as 100cm and the screen display did not show any visual feedback. Only voice was played through the telepresence robot using the Wizard of Oz methodology. Participants received instructions of the study 2 and were instructed to ask questions or requests to the conversational agent. After each question or request by the participants, the experimenter played the recorded voices that answered the question. This process made participants feel like they were directly communicating with the prototype. The experimenter encouraged participants to complete the survey by focusing only on the voices they heard.

The results of the study were analyzed with two-way ANOVA to see the different values of personality perceptions depending on two variables: gender of voice and verbal cues. In addition, to compare more detailed differences between each group, the Tukey HSD post-hoc test was conducted. More detailed results are described in the following result section.

4.4. Results

4.4.1. Result 1: Pitch levels and gender of voices

Two-way ANOVA was applied to compare the perceived personalities depending on the different pitch levels and gender of voices. In the data analysis for this part, a total of 12 participants' data was analyzed, and each participant was exposed to six different voices: gender of voices (female/male) x pitch levels (high/moderate/low). There were no missing values. Among five personalities, conscientiousness showed differences depending on different pitch levels and gender of voices. In two-way ANOVA results concerning the impacts of both pitch level and gender of voices on personality perceptions, the perception of conscientiousness showed significant differences depending on the gender of voices. The detailed results of two-way ANOVA for conscientiousness are shown in Table 8.

Table 8 Results of two-way ANOVA

Personality	Verbal cues		Mean (SD)	df	F value	P value
Conscientiousness	Pitch level	High	5.16 (1.20)	2	0.04	0.96
		Moderate	5.25 (1.19)			
		Low	5.25 (1.25)			

	Gender	Female	5.58 (0.98)	1	6.79	0.01
		Male	4.86 (1.30)	2	0.79	0.46

Based on the two-way ANOVA results, to examine more detailed differences between genders of voices in the perception of conscientiousness, Tukey HSD post-hoc test was conducted. The results showed that female voice (M= 5.58) was more likely to be perceived as more conscientious than the male voice (M= 4.86). Each voice with different pitch levels did not show significant differences depending on the perceptions of personalities.

Table 9 Results of Tukey HSD post-hoc test for pitch

Personality	Verbal cues	Family-wise comparisons	differe nce	95% Confidence level		p value
				lower	upper	
Conscientio usness	Gender	Female voice - male voice	0.72	0.17	1.27	0.01*

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

4.4.2. Result 2: Emotionality and Gender of voices

Two-way ANOVA was conducted to compare the mean values of personality perception depending on different levels of emotionality and gender of voices. In the data analysis for this part, a total of 12 participants' data was analyzed, and each participant was exposed to four different voices: gender of voices (female/male) x emotionality (high/low). There were no missing values. The results of two-way ANOVA demonstrate that the perception of openness and extraversion showed significant differences depending on the level of emotionality of voices. Other personalities including agreeableness, neuroticism, and conscientiousness did not show significant differences in personality perception depending on different levels of emotionality and gender of voices. Two-way ANOVA results for openness and extraversion are shown in Table 10.

Table 10 Results of two-way ANOVA for emotionality

Personality	Verbal cues		Mean (SD)	df	F value	P value
Extraversion	Emotionality	High	6.27 (0.69)	1	140.47	0.00
		Low	3.60 (0.92)			

	Gender	Female	5.14 (1.47)	1	3.42	0.07
		Male	4.72 (1.66)			
	Gender * Emotion ality	Female * high	6.33 (0.61)	1	1.68	0.20
		Male * high	6.20 (0.78)			
Female * low		3.95 (1.03)				
Male * low		3.25 (0.65)				
Openness	Emotion ality	High	4.83 (1.35)	1	7.16	0.01
		Low	3.58 (1.78)			
	Gender	Female	4.22 (1.76)	1	0.00	0.92
		Male	4.18 (1.65)			
	Gender * Emotion ality	Female * high	4.95 (1.43)	1	0.19	0.65
		Male * high	4.70 (1.32)			
		Female * low	3.5 (1.8)			
		Male * low	3.66 (1.83)			

Conscientiousness	Emotionality	High	4.17 (1.22)	1	8.96	0.00
		Low	5.21 (1.17)			
	Gender	Female	4.63 (1.30)	1	0.13	0.72
		Male	4.75 (1.31)			
	Gender * Emotionality	Female * high	4.29 (1.23)	1	1.16	0.29
		Male * high	4.04 (1.23)			
		Female * low	4.96 (1.32)			
		Male * low	5.46 (0.99)			

Based on the result of two-way ANOVA, Tukey HSD post-hoc test was conducted to see the differences of personality perceptions depending on the level of emotionality more specifically. The results of Tukey HSD post-hoc demonstrated that high emotionality (M= 6.27) was perceived as more extraverted than low emotionality (M= 3.60), regardless of voice gender. Male voice with low emotionality (M=

5.46) was more perceived as conscientious than male voice with high emotionality (M= 4.04). In addition, the voice with high emotionality (M= 4.83) was more perceived as openness than voice with low emotionality (M= 3.58) regardless of gender. The result shows that only the level of emotionality was influential, not the gender of the voices. The results of Tukey post-hoc test for perceptions of three personalities (extraversion, openness, and conscientiousness) depending on different emotionality and gender of voices are shown in Table 11.

Table 11 Results of Tukey HSD post-hoc test of emotionality

Personality	Verbal cues	Family-wise comparisons	Difference	95% confidence level		P value
				lower	upper	
Extraversion	Emotionality	Low emotionality – High emotionality	-2.67	-3.14	-2.19	.00***
	Gender & emotionality	Female high emotionality – Male high emotionality	0.125	-0.73	0.98	.98
		Male low emotionality – Male high emotionality	-2.96	-3.82	-2.10	.00***
		Female low emotionality – Male high	-2.25	-3.11	-.139	.00 ***

		emotionality				
		Male low emotionality – Female high emotionality	-3.08	-3.94	-2.22	.00***
		Female low emotionality– Female high emotionality	-2.38	-3.23	-1.52	.00***
		Female low emotionality – Male low emotionality	0.71	-0.15	1.57	.14
Openness	Emotionality	Low emotionality – High emotionality	-1.25	-2.17	-0.33	.00***
Conscientiousness	Emotionality	Low emotionality – High emotionality	1.04	0.35	1.74	.00***
	Gender & emotionality	Female high emotionality – Male high emotionality	0.25	-0.16	1.56	.96
		Male low emotionality – Male high emotionality	1.42	0.10	2.73	.03*
		Female low emotionality – Male high emotionality	0.92	-0.40	2.23	.26
		Male low emotionality – Female high emotionality	1.17	-0.15	2.48	.10
		Female low emotionality -Female high emotionality	0.67	-0.65	1.98	.53
		Female low emotionality – Male low emotionality	-0.50	-1.81	0.81	.74

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

4.4.3. Result 3: Wordiness and gender of voices

Two-way ANOVA was conducted to compare the mean values of personality perception depending on different levels of emotionality and gender of voices. In the data analysis for this part, a total of 12 participants' data were analyzed, and each participant was exposed to four different verbal cues: gender of voices (female/male) x wordiness (high/low). There were no missing values. The results of two-way ANOVA demonstrated that perception of neuroticism differed depending on the level of wordiness. The mean values of other personalities did not show the significant differences depending on the wordiness and gender of voices. The detailed results are given in Table 12.

Table 12 Results of two-way ANOVA for wordiness

Personalities	Verbal cues		Mean (SD)	df	F value	P value
Neuroticism	Wordiness	High	2.67 (1.07)	1	3.2	.087
		Low	3.33 (0.71)			

Tukey post-hoc test was conducted to examine detailed differences in values of personality perceptions depending on the wordiness. According to the post-hoc test result, less wordy speech style (M= 3.33) was more likely to be perceived as neurotic than wordy speech style, regardless of gender of voices (Table 13).

Table 13 Results of Tukey post-hoc test for wordiness

Personality	Verbal cues	Family-wise comparisons	difference	95% Confidence level		p value
				lower	upper	
Neuroticism	Wordiness	Low wordiness-High wordiness	0.66	-0.1	1.43	.00***

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

4.4.4. Result 4: Speed and gender of voices

To compare the different mean values of personalities depending on different speech speed and gender of voices, two-way ANOVA was conducted. In the data analysis for this part, a total of 12 participants' data were analyzed, and each participant was exposed to six different verbal cues: gender of voices (female/male) x speed (fast/moderate/slow). There were no missing values. Among the five

personalities, four personalities (agreeableness, conscientiousness, extraversion, and neuroticism) demonstrated the differences depending on gender and speech speed. Different speech speed and gender of voices did not show significant differences in the perception of openness. The result of two-way ANOVA for agreeableness, conscientiousness, extraversion, and neuroticism are shown in Table 14.

Table 14 Results of two-way ANOVA for speech speed

Personalities	Verbal cues		Mean (SD)	df	F value	P value
Agreeableness	Voice Gender	Male	4.92 (1.54)	1	4.53	.04
		Female	5.63 (1.20)			
	Speech Speed	Fast	5.15 (1.66)	1	0.4	.67
		Moderate	5.48 (1.19)			
		Slow	5.19 (1.39)			
Conscientiousness	Voice Gender	Male	4.40 (1.67)	1	13.45	.00
		Female	5.54 (1.06)			
	Speech Speed	Fast	5.00 (1.56)	2	1.54	.22
		Moderate	4.62			

			(1.76)			
		Slow	5.29 (1.08)			
	Gender * Speed	Male * Fast	4.79 (1.65)	2	4.81	.01
		Male * slow	3.37 (1.53)			
		Male * Moderate	5.04 (1.38)			
		Female *fast	5.20 (1.49)			
		Female *slow	5.87 (0.85)			
		Female* Moderate	5.54 (0.62)			
		Voice Gender	Male			
	Female		3.75 (1.28)			
	Speech Speed	Fast	4.12 (0.81)	1	5.00	.01
		Moderate	3.16 (1.46)			
		Slow	3.18 (1.23)			
Extraversion	Gender * Speed	Male * Fast	3.96 (0.81)	2	0.15	.86
		Male * slow	2.83 (1.20)			
		Male * Moderate	2.91 (1.31)			
		Female *fast	4.29 (0.81)			
		Female *slow	3.54 (1.21)			
		Female* Moderate	3.41 (1.62)			

Neuroticism	Voice Gender	Male	3.09 (1.30)	1	0.15	.70
		Female	2.98 (1.31)			
	Speech Speed	Fast	3.52 (1.49)	1	2.8	.06
		Moderate	2.72 (0.97)			
		Slow	2.87 (1.29)			
		Gender * Speed	Male *	3.50	2	0.08
Fast			(1.31)			
Male *			3.00			
slow			(1.55)			
Male *			2.79			
Moderate			(1.01)			
Female *fast			3.54 (1.72)			
Female *slow	2.75 (1.03)					
Female*	2.67 (0.98)	Moderate				

Tukey HSD post-hoc test was conducted to compare the mean values of personalities depending on voices with different gender and speech speed. In case of extraversion, participants perceived female voice speaking fast (M= 3.54) as extraverted. In case of conscientiousness, participants perceived female voice speaking with slow (M= 5.87) and moderate (M= 5.54) speeds as more conscientious than female voice speaking fast (M= 5.20). In case of neuroticism, male

voice speaking fast (M= 3.50) was perceived as neurotic. The results of Tukey post-hoc test of conscientiousness, extraversion, and neuroticism that showed the significant values are shown in Table 15.

Table 15 Results of Tukey post-hoc test for speech speed

Personal ity	Verbal cues	Family-wise Comparison	Differ ence	95% confidence level		p- value
				lower	upper	
Conscie ntiousne ss	Gender	Female voice – Male voice	1.14	0.48	1.8	.00***
	Gender & speech speed	Male fast – Female fast	-1.83	-3.41	-0.25	.01
		Female slow – Male slow	2.5	0.92	4.08	.00***
		Male moderate – Male slow	1.67	0.09	3.25	.03
		Female moderate	2.17	0.59	3.75	.00**

		- Male slow				
Extraversion	Gender	Female voice - Male voice	0.51	-0.07	1.1	.08
	Speech speed	Slow speed - Fast speed	-0.94	-1.77	-0.11	.02*
		Moderate speed - Fast speed	-0.96	-1.79	-0.13	.02*
		Moderate speed - Slow speed	-0.02	-0.85	0.81	1.00
	Gender & speech speed	Male slow - Female fast	-1.46	-2.89	-0.03	.04*
		Male slow - Male fast	-1.38	-2.81	0.06	.07
	Neuroticism	Gender	Female voice - Male voice	-0.81	-1.39	-0.22
Speech speed		Slow speed - Fast speed	-0.65	-1.53	0.24	.19
		Moderate speed	-0.79	-1.67	0.09	.09

		- Fast speed				
		Moderate speed - Fast speed	-0.15	-1.03	0.74	.92

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

4.4.5. Result 5: Questioning and gender of voices

Two-way ANOVA was conducted to compare the different mean values for the five personalities depending on different gender of voices and whether the conversational agent’s speech includes the questionnaires. In the data analysis for this part, a total of 12 participants’ data were analyzed, and each participant was exposed to four different verbal cues: gender of voices (female/male) x questioning (high/low). There were no missing values. Only extraversion and openness demonstrated significant differences between different usage of questionnaires and gender of voices. The result of two-way ANOVA is shown in Table 16.

Table 16 Results of two-way ANOVA for questionnaires

Personalities	Verbal cues	Mean (SD)	df	F value	P value
---------------	-------------	-----------	----	---------	---------

Extraversion	Question	High	3.85 (1.57)	1	4.15	0.04	
		Low	3.10 (0.98)				
	Gender	Female	3.10 (1.38)	1	4.15	0.04	
		Male	3.85 (1.24)				
	Gender * Question	Female * high	3.08 (1.54)	1	4.62	0.03	
		Male * high	4.62 (1.22)				
		Female * low	3.12 (1.26)				
		Male * low	3.08 (0.66)				
	Openness	Question	High	4.41 (1.41)	1	0.34	0.04
			Low	4.18 (1.28)			
Gender		Female	4.70 (1.16)	1	4.37	0.55	
		Male	3.89 (1.39)				
Gender * Question		Female * high	4.79 (1.26)	1	0.02	0.87	
		Male * high	4.04 (1.49)				
		Female * low	4.62 (1.11)				
		Male * low	3.75 (1.33)				

According to the Tukey post-hoc test results, in case of

extraversion, the conversational agent with male voice that questions a lot is perceived more as the most extraverted (M= 4.62). Other verbal cues such as female voice that questions a lot (M= 3.08), male voice that does not question a lot (M= 3.08), and female voice not using questions (M= 3.12) were less perceived as extraversion. In case of openness, people perceived female voice (M= 4.70) more as openness than male voice (M= 3.89). Results of comparison in extraversion and openness are shown in Table 17.

Table 17 Results of Tukey post-hoc test for questioning

Personality	Verbal cues	Family-wise comparisons	difference	95% Confidence level		p value
				lower	upper	
Extraversion	Gender * Question	Female * high - Male * high	-0.15	-2.87	-0.21	0.01*
		Male * low - Male * high	-0.15	-2.87	-0.21	0.01*
		Female * low - Male * high	-0.15	-2.82	-0.17	0.02*
		Male * low - Female * high	-0.44	-1.32	1.32	1*
		Female * low - Female *	-0.47	-1.28	1.37	0.99

		high				
		Male * low - Female * low	-0.41	-1.28	1.37	0.99
Openness	Gender	Female - male	0.81	0.06	1.56	0.03*

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

4.5. Overall results

Based on two-way ANOVA results and Tukey post-hoc tests, the final results of study 2 for each personality are as follows (Table 18).

Table 18 Overall results of study 2

Personality Verbal cues	Conscientious	Extraversion	Openness	Neuroticism
Emotionality	Male low emotionality	High emotionality	High emotionality	-
Speed	Female slow, moderate speed	Female fast speed	-	Male fast speed
Question	-	Male voice questions a lot	Female voice	-
Wordiness	-	-	-	Low wordiness

Pitch	Female voice	-	-	-
-------	--------------	---	---	---

Conscientiousness is associated with being careful, diligent, painstaking, and thorough. Male voice with low emotionality is perceived more as conscientious than voices with high emotionality. Female voice speaking with slow and moderate speaking speed is more likely to be perceived as conscientious than fast speaking speed.

Extraversion is associated with being sociable, gregarious, assertive, talkative, actively ambitious, and initiative. Voice with high emotionality was likely to be perceived as extraverted personality regardless of the gender of voices. Fast speaking female voice was perceived more as active and sociable than voices with moderate, slow speaking speed.

Openness is associated with being creative, artistically sensitive, and intelligent. Voice with high emotionality regardless of gender of voices was more likely to be perceived as openness than voice with low emotionality.

Neuroticism is also known as emotionally unstable and is related to the degree of negative emotion, such as being moody and tense. Regardless of gender of voices, not wordy speech style was

perceived as more neurotic than talkative speech style. Male voice speaking fast was more likely to be perceived as neurotic than other speech speed conditions.

Agreeableness was excluded in the table because verbal cues condition did not show a significant value.

4.6. Discussion

Perceptions of the conversational agent's personalities differed by gender of voices. Fast speaking speed was perceived differently depending on the gender of voices. When male voice speaks fast, it was perceived as emotionally unstable and nervous, but female voice speaking fast was perceived as sociable and outgoing. Sociable and outgoing personalities are considered more positive personalities than unstable and nervous personalities (McCrae & Costa, 1997) Even though the same verbal cue was used, the personality perception differed depending on gender of voices. Especially, people tend to perceive female voice as more positive than male voice. This result is highly related with the previous study findings that people prefer female extraverted voices for assistive social robots for elders (Chang et al., 2018).

The study results show that female voices were more considered as having personality than male voices. In other words, when only considering the gender of voices, female voices demonstrated meaningful values while male voice did not. Female voice was perceived as creative, calm, and outgoing without verbal cue conditions. However, male voice was perceived as certain personalities only when combined with other verbal cues. This could imply that participants tend to react sensitively to female voices.

Gender stereotypes exist in how the agent's personality is perceived. The present study results demonstrate that people apply gender stereotypes that typically characterize the human-human social cognitive process to the interfaces, which supports previous findings (Nass et al., 2013; Eyssel & Hegel, 2012). This induces the ethical question about the designs of robots, specifically, whether the gender stereotypes must be included or excluded when designing computer systems and machines. Some researchers believe that stereotypes and prejudices must be changed, and that a new approach is needed to address the gender issue (Cassell, 1986). Others suggest that building gender-congruent robots is better, because it is the simplest way to meet the users' expectations (Nass et al., 2013).

The interactions between conversational agents and humans demonstrated similar patterns of perception with human–human interactions. Similar with the conversational rules in human–human interactions, using diverse and expressive verbal cues were suitable for positive personalities. When using diverse verbal cues, the conversational agent is more likely to be perceived as having positive and active personalities. For instance, male voice questioning a lot is perceived as extraversion, while less wordy speech style is considered as neurotic.

The study results are similar with the previous study in that extroverts are more talkative than introverts in human–human interactions (Dewaele & Furnham, 1999). In addition, the results support the finding that people enjoy chatbots with a distinct personality (Jain et al., 2018).

4.7. Limitations

The experiment was conducted with scripts that had been already written and recorded in advance. Since the study was not conducted in a real situation concerning specific contexts, future studies must consider the context of real-time interactions between people and the

social interfaces. For future studies, rather than strictly focusing on dedicated scripts and scenarios, more real-situational approaches are needed (Breazeal, 2004).

5. Study 3

5.1. Overview

Certain personalities are considered better suited than others for particular tasks (Barrick & Mount, 1991; McCrae & Costa, 1997). When an individual has certain personality traits that are suitable for performing the task, this could even lead to higher chances of getting the job (Glick, 1991). In addition, both personality and gender are strong indicators of successful careers and job performances (Alter & Seta, 2005). Results similar to those in human–human communication studies were demonstrated in human–robot interaction studies in that people preferred a robot with matching gender-occupational roles and personality-occupational roles (Tay et al., 2014).

The aim of study 3 is to examine the match between the conversational agent’s personalities and tasks it performs. It is the extended study based on a conceptual study (Lee et al., 2018), observing the expected personality for expected robot tasks. It is different from previous studies in that the personality expression was conducted based on the results of study 1 and 2. In addition, the personality was expressed with diverse visual feedback and verbal cues.

Tasks performed by the conversational agent was manipulated with the contents it delivers, also based on the expected tasks. The main research question is that ***“RQ 3. Which personality do users prefer for which tasks performed by the conversational agent?”***

5.2. Method

5.2.1. Experimental Materials

A total of 16 stimuli (4 personalities * 4 task types) were designed in the format of video. Each stimulus was designed with visual feedback (results of study 1) and verbal cues (results of study 2).

The manipulation check (section 5.2.2.) for the combined visual feedback and verbal cues were conducted because experiments were conducted separtely. The results of study 2 cannot be entirely applied as the experimental materials of study 3 because both female and male voices were statistically significant for conscientiousness and extraversion. For verbal cues, p-values and mean values of study 2 results were compared and gender of voices was selected.

Similar with study 2 (section 4.3.1.), the female and male voices from Amazon Polly were set as default value in study 3. Different

conditions of verbal cues were adjusted with the programs such as Audacity and Oddcast. Experimental materials of study 2 with detailed verbal cues and visual feedback are mentioned in Table 19.

Table 19 The experimental materials of study 3

Personality Feedback	Extraversion	Conscientiousness	Openness	Neuroticism
Visual feedback	Fast motion	Moderate motion	Moderate motion	Slow motion
Verbal cues	Fast speed, high emotionality, questioning a lot	Low emotionality, slow speed	High emotionality	Fast speed, low wordiness
Gender of voices	Male	Female	Female	Male

To express extraversion, visual feedback with fast motion was used. For verbal cues, male voice that speaks fast, expresses emotionally, and questions a lot was used for the experiment. Male voice was selected because mean and p-values for male voice questioning a lot condition were statistically more significant than female voice speaking fast.

To express conscientiousness, for visual feedback, moderate motion was used. For verbal cues, female voice that speaks slow with low emotionality was used. Female voice was selected because of

female voice talking slowly demonstrated statistically more significant mean and p-values than male voice with low emotionality.

To express openness, for visual feedback, moderate motion was used. For verbal cues, female voice with high emotionality was used. To express neuroticism, male voice that speaks fast but does not talk much was used.

The task types were differentiated depending on the contents that the experimental material delivers, because the aim of the study was to find the preferred personality based on task types. The scenarios were written based on four categories of tasks. The study examined people's expectations and preferences for a robot's personality based on the tasks the robot is performing with interviews and surveys (Lee et al., 2018). Four categories of tasks were expected to be completed by social robots.

Scenarios for each task were as follows. For social tasks, scripts of conversational agents were game playing, rapping, compliment, introduction, and time-killing joke. For service task, the scenarios were giving directions, weather forecast, airplane ticket suggestion, coffee shop order, and restaurant serving. For physical tasks, the scenarios were cleaning the room, driving, taking an order, package delivery, and

home security. For office tasks, scenarios were admission counseling, health condition, printing, typing, and writing documents.

Each scenario was recorded with different verbal cues for each personality referring to the results of study 1, 2, 3. For instance, in the case of extraversion, the video was made by combining visual feedback with fast motion saying a time-killing joke with the voice with high pitch female, high emotionality, and fast speaking speed.

5.2.2. Manipulation check

Participants evaluated the combinations of visual feedback and verbal cues they had exposed with on the following bipolar scales ($\alpha = .83$, $M = 3.9$, $SD = 1.6$): “reserved (1)-- sociable, outgoing (7)”, “lazy – does a thorough job”, “nervous-- relaxed”, “few artistic interests – active imagination” (Costa & McCrae, 2017).

One-sample t-test was conducted respectively for each personality condition. Male voice with high emotionality that speaking fast and questions a lot ($M = 4.6$, $SD = 1.14$) was rated significantly high in bipolar scales measuring extraversion. Female voice with high emotionality was rated significantly high in bipolar scales measuring openness ($M = 5.0$, $SD = 1.43$). Male voice with fast speaking but less

wordiness was rated high in neuroticism questionnaire (M= 4.57, SD= 1.24). Female voice with no emotionality that speaking slowly was rated high in bipolar scales of conscientiousness (M= 4.1, SD= 1.4). Four personalities were rated different from the scale midpoint (4) in the intended direction.

5.2.3. Experimental Setting

A total of 68 participants (17 participants * 4 tasks) were recruited for the experiment; 17 participants were exposed to each experimental material, which is the combination of verbal cues and visual feedback with different tasks. Since it is difficult to implement all tasks in real situation with a telepresence robot, one-sentence-length brief explanations related with the tasks were given to participants before being exposed to experimental material. For instance, in case of a service task such as cleaning the room, the researcher briefly explained the context of the service task.

After being exposed to the experimental material, participants filled out the survey questionnaire that was based on the Godspeed scale (Bartneck et al., 2009). The Godspeed scale questionnaire was chosen because it is one of the standardized measurement tools for

human–robot interaction. Among five key concepts of Godspeed scale questionnaires, the study adopted two concepts (likeability and perceived intelligence) because these two concepts best match with the aim study. The full questionnaires are written in Appendix 2.

The data of the 68 participants were analyzed with one-way ANOVA to compare the likeability and perceived intelligence for each personality and task combination. In addition, more detailed differences between likeability and perceived intelligence were compared with Tukey HSD post-hoc test.

5.3. Results

5.3.1. Office task

According to the one-way ANOVA result of the office task (Table 20), both the likeability and perceived intelligence demonstrated significant differences depending on the personalities. The likeability of the conversational agent ($\alpha = 0.9$) that performs the office task differed depending on the personality the conversational agent expresses ($F(3) = 5.52$, $p < 0.01$). In addition, the perceived intelligence ($\alpha = 0.87$) differed depending on the personality the conversational agent

expresses ($F(3) = 7.89, p < 0.01$). More detailed results for one-way ANOVA are demonstrated in Table 20.

Table 20 Results of one-way ANOVA for office task

	Groups	Mean	SD	df	F-value	P-value
Likeability	Extraversion	3.90	1.30	3	5.52	.000
	Openness	5.34	1.28			
	Conscientiousness	3.94	1.26			
	Neuroticism	4.47	1.24			
Perceived Intelligence	Extraversion	3.91	1.46	3	7.89	.000
	Openness	5.88	1.42			
	Conscientiousness	4.32	1.39			
	Neuroticism	4.5	1.36			

Tukey HSD post-hoc test was conducted to compare the detailed differences of likeability and perceived intelligence when conducting the office task with different personalities. Participants most liked the conversational agent performing office task when having a creative personality ($M = 5.34$), especially compared with extraversion ($M = 3.90$) and conscientiousness ($M = 3.94$). In addition, participants thought the creative ($M = 5.88$) conversational agent as more intelligent than active ($M = 3.91$), neurotic ($M = 4.50$), and conscientious ($M = 4.32$) conversational agents. More detailed results of Tukey HSD post-

hoc test are demonstrated in Table 21.

Table 21 Results of Tukey HSD post-hoc test for office task

	Pair-wise comparisons	Mean differences	Lower	Upper	P-value
Likeability	Openness-Extraversion	1.43	0.37	2.49	0.00**
	Conscientiousness-Extraversion	0.03	-1.03	1.09	1
	Neuroticism - Extraversion	0.56	-0.5	1.62	0.51
	Conscientiousness-Openness	-1.4	-2.46	-0.34	0.00**
	Neuroticism-openness	-0.87	-1.93	0.19	0.14
	Neuroticism - Conscientiousness	0.53	-0.53	1.59	0.56
Perceived Intelligence	Openness-Extraversion	1.97	0.83	3.11	0.00***
	Conscientiousness-Extraversion	0.41	-0.72	1.55	0.77
	Neuroticism - Extraversion	0.59	-0.55	1.72	0.52
	Conscientiousness-openness	-1.56	-2.69	-0.42	0.00**
	Neuroticism-openness	-1.38	-2.52	-0.25	0.01*
	Neuroticism - Conscientiousness	0.18	-0.96	1.31	0.98

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

5.3.2. Social task

According to the one-way ANOVA result of the office task (Table 20), both the likeability and perceived intelligence demonstrated significant differences depending on the personalities. The likeability ($\alpha = 0.9$) toward the conversational agent that performs the office task differed according to the personality the conversational agent expresses ($F(3) = 4.02, p < 0.05$). The perceived intelligence ($\alpha = 0.87$) toward the conversational agent conducting the social task differed according to the personality the conversational agent expresses ($F(3) = 3.15, p < 0.05$). More detailed results of one-way ANOVA for social task are demonstrated in Table 22.

Table 22 Results of one-way ANOVA for social task

	Groups	Mean	SD	df	F value	P value
Likeability	Extraversion	5.54	1.54	3	4.02	0.01 *
	Openness	4.75	1.54			
	Conscientiousness	4.34	1.54			
	Neuroticism	3.92	1.54			
Perceived Intelligence	Extraversion	5.05	1.56	3	3.15	0.03 *
	Openness	4.82	1.55			
	Conscientiousness	4.41	1.56			
	Neuroticism	3.62	1.56			

For more detailed comparisons, Tukey HSD post-hoc test was conducted. For the conversational agent conducting the social task, personality of extraversion (M= 5.54) was more liked than neuroticism (M= 3.92). Regarding perceived intelligence ($\alpha = 0.87$), extraverted (M= 5.05) conversational agent was more perceived as intelligent than neurotic (M= 3.62) conversational agent when conducting the social task. More detailed results of Tukey HSD post-hoc test are described in Table 23.

Table 23 Results of Tukey HSD post-hoc test for social task

		Differences	Lower	Upper	P-value
Likeability	Openness- Extraversion	-0.79	-2.08	0.49	0.37
	Conscientiousness- Extraversion	-1.21	-2.49	0.08	0.07
	Neuroticism - Extraversion	-1.62	-2.91	-0.34	0.01 **
	Conscientiousness- Openness	-0.41	-1.7	0.88	0.83
	Neuroticism- Openness	-0.83	-2.12	0.46	0.33
	Neuroticism - Conscientiousness	-0.42	-1.71	0.87	0.83
Perceived intelligence	Openness- Extraversion	-0.24	-1.57	1.09	0.97
	Conscientiousness- Extraversion	-0.65	-1.98	0.68	0.58
	Neuroticism - Extraversion	-1.44	-2.77	-0.11	0.03 *

	Conscientiousness- Openness	-0.41	-1.74	0.92	0.85
	Neuroticism- Openness	-1.21	-2.54	0.12	0.09
	Neuroticism - Conscientiousness	-0.79	-2.12	0.54	0.4

Signif. codes: $p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *

5.3.3. Service task

According to the one-way ANOVA result for the service task, the likeability towards the conversational agent conducting the service task demonstrated the significant differences depending on its personalities ($F(3) = 3.89$). The perceived intelligence demonstrated significant differences depending on the conversational agents' personalities ($F(3) = 4.20$) when conducting the service task. More detailed results of one-way ANOVA for the service task are demonstrated in Table 24.

Table 24 Results of One-way ANOVA for service task

	Groups	Mean	SD	df	F-value	P-value
Likeability	Extraversion	4.41	1.45	3	3.89	0.012 *
	Openness	4.77	1.41			

	Conscientiousness	4.11	1.4			
	Neuroticism	3.27	1.4			
Perceived Intelligence	Extraversion	4.38	1.52	3	4.2	0.00 **
	Openness	4.77	1.48			
	Conscientiousness	4.08	1.46			
	Neuroticism	3.14	1.46			

To compare more detailed differences between users' perceptions including perceived intelligence and likeability, Tukey HSD post-hoc test was conducted (Table 25). Creative and intellectual (M= 4.77) conversational agent conducting the service task was more liked than deliberate and careful agent (4.08). In addition, creative (M= 4.77) conversational agent conducting the service task was more perceived as intelligent than outgoing (M= 4.41), careful (M= 4.11), and anxious (M= 3.27) conversational agents. More detailed results of Tukey HSD post-hoc test for service task are given in Table 25.

Table 25 Results of Tukey HSD post-hoc test for service task

	Pair-wise comparisons	Mean Differences	Lower	Upper	P-value
--	-----------------------	------------------	-------	-------	---------

Likeability	Openness- Extraversion	1.43	0.37	2.49	0.00**
	Conscientiousness- Extraversion	0.03	-1.03	1.09	0.99
	Neuroticism - Extraversion	0.56	-0.5	1.62	-0.51
	Conscientiousness- Openness	-1.4	-2.46	-0.33	0.00**
	neuroticism- Openness	-0.87	-1.93	0.19	0.14
	Neuroticism - Conscientiousness	0.53	-0.53	1.59	0.56
Perceived Intelligence	Openness- Extraversion	1.97	-0.83	3.11	0.00***
	Conscientiousness- Extraversion	0.41	-0.72	1.55	0.77
	Neuroticism - Extraversion	0.59	-0.55	1.72	0.52
	Conscientiousness- Openness	-1.56	-2.69	-0.42	0.00***
	Neuroticism- Openness	-1.39	-2.52	-0.25	0.01 *
	Neuroticism - Conscientiousness	0.18	-0.96	1.31	0.98

Signif. codes: $p < 0.001$ ***, $p < 0.01$ **, $p < 0.05$ *

5.3.4. Physical task

According to the results of one-way ANOVA (Table 26) for the conversational agent conducting the physical task, likeability demonstrated significantly different values depending on different personalities ($F(3) = 6.62$, $p < 0.001$). However, the perceived

intelligence of the conversational agent conducting the physical task did not show significantly different values depending on different personalities ($F(3) = 1.77$). More detailed results of one-way ANOVA are given in Table 26.

Table 26 Results of One-way ANOVA for physical task

	Groups	Mean	SD	df	F-value	P-value
Likeability	Extraversion	4.20	1.4	3	6.62	0.00***
	Openness	5.25	1.38			
	Conscientiousness	3.44	1.38			
	Neuroticism	3.91	1.34			
Perceived Intelligence	Extraversion	4.38	1.42	3	1.77	0.16
	Openness	5.14	1.39			
	Conscientiousness	4.12	1.39			
	Neuroticism	4.38	1.35			

Tukey HSD post-hoc test was conducted to analyze more detailed differences (Table 27). Creative conversational agent ($M = 5.25$) was more liked than outgoing ($M = 4.20$) and conscientious ($M = 3.44$) conversational agent when conducting the physical task.

Conversational agent with creative personality (M= 5.14) was more perceived as intelligent than sociable (M= 4.38), conscientious (M= 4.12), and neurotic (M= 4.38) conversational agent when conducting the physical task. More detailed results of the Tukey HSD post-hoc test for the physical task are demonstrated in Table 27.

Table 27 Results of Tukey HSD post-hoc test for physical task

	Pair-wise comparisons	Mean Differences	Lower	Upper	P-value
Likeability	Openness-Extraversion	1.43	0.37	2.49	0.00**
	Conscientiousness-Extraversion	0.03	-1.03	1.09	1
	Neuroticism - Extraversion	0.56	-0.5	1.62	0.51
	Conscientiousness-Openness	-1.4	-2.46	-0.34	0.00**
	neuroticism-Openness	-0.87	-1.93	0.19	0.14
	Neuroticism - Conscientiousness	0.53	-0.53	1.59	0.56
Perceived Intelligence	Openness-Extraversion	1.97	0.83	3.11	0.00***
	Conscientiousness-Extraversion	0.41	-0.72	1.55	0.77
	Neuroticism - Extraversion	0.59	-0.55	1.72	0.52
	Conscientiousness-Openness	-1.56	-2.69	-0.42	0.00**
	Neuroticism-Openness	-1.38	-2.52	-0.25	0.01*

	Neuroticism - Conscientiousness	0.18	-0.96	1.31	0.98
--	------------------------------------	------	-------	------	------

Signif. codes: p< 0.001 ***, p<0.01 **, p<0.05 *

5.4. Discussion

Openness and extraversion were two personalities that were liked by users and perceived as the most intelligent when the conversational agents conduct the tasks.

Creative, imaginative, and intellectual personalities were mostly liked and perceived as intelligent in service, office, and physical tasks compared with other personalities. This is highly related with and also supports the findings from a previous study that female and extroverted voices were preferred by most participants (Chang et al., 2018) because openness was expressed with female voice with high emotionality. In addition, users expect the agents to have creative and imaginative personalities.

Exceptionally, when conducting the social task, people expect the conversational agent to be sociable and outgoing, rather than creative. Considering that extraversion was expressed using the visual feedback with fast motion and verbal cues with high emotionality, fast speaking

speed that questions a lot, the study results imply that users expect the conversational agents to be sociable and outgoing when having a small talk.

Results of study 3 imply that users' perceptions of agents are influenced by different types of personality expression cues. This results could be considered with the previous study about chatbot in that people expect chatbot's personality to match with the task (Jain et al., 2018). Meanwhile, the detailed match between personality and tasks was different. People expect a news chatbot to be professional, while a shopping chatbot to be humorous and casual (Jain et al., 2018). Study 3 results demonstrated that intelligent and creative personalities are needed both for news briefing and shopping related tasks. Casual and humorous personalities were rather required for the conversational agent performing small talk. This different match between personalities and tasks sought to be due to different personality expression cues. Chatbot delivers its personalities with written texts, but the conversational agent uses voice and visual feedback to express personalities in this study.

6. Conclusions

This study has explored how the conversational agent's personalities could be expressed. In addition, the preference for and perceived intelligence of conversational agents depending on its personality and the tasks it performs were examined.

The current study provides a novel approach to implement natural feedback of conversational agents with the concept of personality rather than emotion. In addition, rather than focusing on two contrasting personalities, the current study applied wider personality categories. The current study sought to implement more natural feedback by applying more diverse personality expression cues.

Studies 1-1 and 1-2 were conducted to answer the first research question "Can different visual feedback be perceived as different personalities?". The results of Studies 1-1 and 1-2 demonstrated that different motions of visual feedback were highly influential on the perception of personality regardless of color. Visual feedback with different motions could express different personalities. Visual feedback with slow motion was perceived as depressed and anxious while fast motion was perceived as active and sociable. Moderate motion was

perceived as having a creative, agreeable, and imaginative personality. Slow motion was perceived as conscientious and neurotic.

Regarding color, the results of study 1 highlighted the issue of color subjectivity and objectivity. Color was not a decisive factor in the perception of personality. The results support the color subjectivity that colors cannot be perceived objectively. The results also support that fast motions are appropriate for expressing positive personality traits such as agreeableness and extraversion. Meanwhile, slow motions are personality traits that are usually perceived as negative. Fast and active visual feedback is suitable in the design of conversational agents with distinct and positive personalities.

Study 2 was conducted to answer the second research question “Can different verbal cues be perceived as different personalities?”. The results of study 2 demonstrate that different verbal cues are perceived as different personalities. Regarding conscientiousness, the male voice with low emotionality and female voice with slow and moderate speed were statistically significant. A voice with high emotionality, female voice with fast speech speed, and male voice that questions a lot were perceived as extraverts. Female voice with high emotionality was likely to be perceived as openness while male voice

speaking fast with low wordiness was perceived as neurotic.

The results of study 2 indicated that the perception of a conversational agent's personality differed according to the voice's gender. Gender stereotypes exist in how a conversational agent's personality is perceived. Female voices were more likely considered as having a personality than male voices. Participants tended to react sensitively to different verbal cues with a female voice. In addition, diverse and expressive verbal cues are better suited for positive and active personalities.

Study 3 was conducted to answer the third research question "Which personality do users prefer for which tasks performed by the conversational agent?". The results of study 3 demonstrated that users prefer the conversational agents with creative and intelligent personalities and perceive this as intelligent when conducting office, physical, and service tasks. Exceptionally, users preferred and perceived the conversational agent with extraversion as intelligent when performing the social task.

The results of study 3 imply that users' perceptions of agents are influenced by different types of personality expression cues. Unlike previous studies about chatbot, intelligent and creative personalities

were needed for office tasks such as shopping and news briefings, while casual and humorous personalities were needed for casual conversations. Chatbot delivers its personalities with written texts, but the conversational agent uses voice and visual feedback to express personalities in this study.

The overall results of the conversational agents could be applied to diverse interfaces designed with smart displays such as artificial intelligent speakers, social robots, cars, and IoT environments. In addition, the study applied a wider range of personalities rather than focusing on two contrasting personalities. Furthermore, more diverse elements were applied in personality expressions rather than focusing on simple factors.

7. Discussion for overall study

The results of this study are applicable in diverse social and industrial interfaces, such as smart speakers with a screen, appearance-constrained social robots, virtual environments, and autonomous cars. Previous studies have been focused on expressions of internal states with human-like conversational agents, which are not applicable to diverse formats of interfaces (Cassell & Thórisson, 1999). The results of the current study are applicable to interfaces that cannot implement human-like elements such as facial expressions, eye gaze, and gestures.

Artificial intelligence technology platforms are changing. For instance, an increasing number of smart speakers are being launched with screen display platforms⁷. By applying the results of visual feedback with motions, it is possible to express the internal states of the smart speakers with the screen displays.

As the Internet of Things has advanced home applications and settings, voice is a decisive factor to deliver the message and internal states. Accordingly, the results of the verbal cues could be applied in

⁷ <https://assistant.google.com/platforms/displays/>

IoT settings.

The results of personality expressions could be applied in assistive social robots for elders (Wu et al., 2012), stressful driving situations (Paredes et al., 2018) and interpersonal haptic devices (Bailenson et al., 2007). For instance, conscientious personality could be implemented with low pitch, low emotionality, and low-speed male voice (rather than extraverted voices) in stressful driving situations. With haptic devices, slow motion of the visual feedback could be used to express virtual avatars' personalities with touching feedback. Considering that small creative social robots are preferred by elders; high emotionality female voice could be applied for small assistive social robots.

References

- A. H. Munsell. (2018). A pigment Color System and Notation. *The American Journal of Psychology*, 23(2), 236–244.
- Adams, A. T., Costa, J., Jung, M., & Choudhury, T. (2015). Mindless Computing : Designing Technologies to Subtly Influence Behavior
Mindless Computing : Designing Technologies to Subtly Influence Behavior. *UbiComp*, (September).
- Alter, R. J., & Seta, C. E. (2005). Compensation for inconsistencies: The effects of stereotype strength on expectations of applicants' job success and satisfaction. *Sex Roles*, 53(1–2), 79–87.
- Aly, A., & Tapus, A. (2013). A Model for Synthesizing a Combined Verbal and Nonverbal Behavior Based on Personality Traits in Human-Robot Interaction A Model for Synthesizing a Combined Verbal and Nonverbal Behavior Based on Personality Traits in Human A Model for Synthesizing a Comb, 325–332.
- Andrist, S., Mutlu, B., & Tapus, A. (2015). Look Like Me: Matching Robot Personality via Gaze to Increase Motivation. *ACM Conference on Human Factors in Computing Systems (CHI)*, 2(1), 3603–3612.
- Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37(5), 715–727.

- Bailenson, J. N., Yee, N., Brave, S., Merget, D., & Koslow, D. (2007). Virtual Interpersonal Touch: Expressing and Recognizing Emotions Through Haptic Devices. *Hum.-Comput. Interact.*, 22(3), 325–353.
- Barrick, M. R., & Mount, M. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*.
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), 71–81.
- Bickmore, T., & Cassell, J. (2005). *Social dialogue with embodied conversational agents. Advances in natural multimodal dialogue systems*.
- Brave, S., & Nass, C. (2003). *Emotion in human–computer interaction*.
- Breazeal, C. (2004). Social interactions in HRI: The robot view. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 34(2), 181–186.
- Brebner, J. (1985). *Personality theory and movement. Individual differences in movement*.
- Byrne, A., & Hilbert, D. R. (2003). Color realism and color science. *Behavioral and Brain Sciences*, 26(1), 3.
- Cassell, J., & Thórisson, K. R. (1999). The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, 13(4–5), 519–538.

- Chang, R. C. S., Lu, H. P., & Yang, P. (2018). Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan. *Computers in Human Behavior, 84*, 194–210.
- Costa, P., & R.Mcrae, R. (2017). *The revised NEO personality inventory (NEO-PI-R)*.
- Dryer, D. C. (1999). Getting personal with computers: how to design personalities for agents. *Applied Artificial Intelligence, 13*(3), 273–295.
- Funakoshi, K., Kobayashi, K., Nakano, M., Yamada, S., Kitamura, Y., & Tsujino, H. (2008). Smoothing human-robot speech interactions by using a blinking-light as subtle expression. In *Proceedings of the 10th international conference on Multimodal interfaces (ICMI)* (p. 293).
- Glick, P. (1991). Trait-based and sex-based discrimination in occupational prestige, occupational salary, and hiring. *Sex Roles, 25*(5–6), 351–378.
- Hancock, J. T., & Dunham, P. J. (2001). Impression Formation in Computer-Mediated Communication Revisited. *Communication Research, 28*(3), 325–347.
- Hyde, J., Carter, E. J., Kiesler, S., & Hodgins, J. K. (2013). Perceptual effects of damped and exaggerated facial motion in animated characters. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*.
- Isbister, K., & Nass, C. (2000). Consistency of personality in interactive

- characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human Computer Studies*, 53(2), 251–267.
- Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and Informing the Design of Chatbots. In *Desing Interactive System* (pp. 895–906).
- Kwak, S. S. (2014). The Impact of the Robot Appearance Types on Social Interaction with a Robot and Service Evaluation of a Robot. *Archives of Design Research*, 1–13.
- Lee, K. M., Peng, W., Jin, S. A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication*, 56(4), 754–772.
- Lee, S., Kim, S., Lee, G., & Lee, J. (2018). Robots in Diverse Contexts : Effects of Robots Tasks on Expected Personality, 169–170.
- Lenclos, J.-P. L. & D. (2004). *Colors of the world*. Norton & Company.
- Liao, Q. V., Hussain, M. M., Chandar, P., Davis, M., Crasso, M., Wang, D., ... Geyer, W. (2018). All Work and no Play ? Conversations with a Question-and-Answer Chatbot in the Wild. *Proc. of CHI*, 1–13.
- Löffler, D., Schmidt, N., & Tscharn, R. (2018). Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound. *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18*, 334–343.

- Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30, 457–500.
- Mairesse, F., & Walker, M. a. (2009). Can Conversational Agents Express Big Five Personality Traits through Language?: Evaluating a Psychologically-Informed Language Generator. *Cambridge & Sheffield, United Kingdom: Cambridge University Engineering Department & Department of Computer Science, University of Sheffield.*, 1–59.
- McCrae, R. R., & Costa, P. T. (1997). Personality Trait Structure as a Human Universal. *American Psychologist*, 52(5), 509–516.
- Nass, C. (2000). Machines and Mindlessness : Social Responses to Computers, 56(1), 81–103.
- Nass, C., & Lee, K. M. (2000). Does computer-generated speech manifest personality? an experimental test of similarity-attraction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '00*, (June), 329–336.
- Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, C. (1995). Can computer personalities be human personalities?, (May 2014), 10–12.
- Neff, M., Wang, Y., Abbott, R., & Walker, M. (2010). Evaluating the Effect of Gesture and Language on Personality Perception in Conversational Agents Evaluating the Effect of Gesture and Language on Personality

- Perception in Conversational Agents, (December).
- Niculescu, A., Van Dijk, B., Nijholt, A., & See, S. L. (2011). The influence of voice pitch on the evaluation of a social robot receptionist. *Proceedings - 2011 International Conference on User Science and Engineering, i-USEr 2011*, 18–23.
- Oliveira, R., Arriaga, P., Alves-Oliveira, P., Correia, F., Petisca, S., & Paiva, A. (2018). Friends or Foes?: Socioemotional Support and Gaze Behaviors in Mixed Groups of Humans and Robots. *ACM/IEEE International Conference on Human-Robot Interaction, Part F1350*,
- Paredes, P. E., Ordoñez, F., Tech, C., & Landay, J. A. (2018). Fast & Furious : Detecting Stress with a Car Steering Wheel.
- Pervin, L. A., & John, O. P. (1999). Handbook of personality.
- Picard. (1999). Affective computing for HCI. *Hci, 1*(August), 829–833.
- Picard, R. W. (1998). Affective Computing. *Trends in Cognitive Sciences*, 2(7), 270.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212.
- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York.
- Schaie, W. (1963). The color pyramid test: a nonverbal technique for

- personality assessment. *Psychological Bulletin*, 40(4), 530–547.
- Sokolova, M., & Fernández-Caballero, A. (2015). A Review on the Role of Color and Light in Affective Computing. *Applied Sciences*, 5(3), 275–293.
- Song, S., & Yamada, S. (2017). Expressing Emotions through Color, Sound, and Vibration with an Appearance-Constrained Social Robot. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, 2–11.
- Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human-robot interaction. *Computers in Human Behavior*, 38, 75–84.
- Terada, K., Yamauchi, A., & Ito, A. (2012). Artificial emotion expression for a robot by dynamic color change. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 314–321.
- Torres, O. E., Cassell, J., & Prevost, S. (1997). Modeling Gaze Behavior as a Function of Discourse Structure. In *First International Workshop on Human-Computer Conversation*.
- Wang, R., Harari, G., Hao, P., Zhou, X., & Campbell, A. T. (2015). SmartGPA : How Smartphones Can Assess and Predict Academic Performance of College Students, (UbiComp), 7–11.
- Warner, S. (1966). On the relation of color and personality. *Journal of Projective Techniques & Personality Assessment*, 30, 512–524.

- Wu, Y. H., Fassert, C., & Rigaud, A. S. (2012). Designing robots for the elderly: Appearance issue and beyond. *Archives of Gerontology and Geriatrics*, *54*(1), 121–126.
- Yee, N., Harris, H., Jabon, M., & Bailenson, J. N. (2011). The Expression of Personality in Virtual Worlds, 5–12.

Appendix 1. Big Five personality questionnaires

How well do the following statements describe the personality of the experimental material? (1= Disagree strongly, 7= Agree strongly)

1. Is reserved
2. Is generally trusting
3. Tends to be lazy
4. Is relaxed, handles stress well
5. Has few artistic interests
6. Is outgoing, sociable
7. Tends to find fault with others
8. Does a thorough job
9. Gets nervous easily
10. Has an active imagination

Scoring the Big Five Inventories-10 scales: (R= item is reversed-scored) Extraversion: 1R, 6; Agreeableness: 2, 7R, Conscientiousness: 3R, 8; Neuroticism: 4R, 9; Openness: 5R; 10

Appendix 2. God speed scale questionnaires

Please rate your impression of the conversational agent on these scales

(1= Disagree strongly, 7= Agree strongly):

Likeability

1. Dislike / Like
2. Unfriendly / Friendly
3. Unkind / Kind
4. Unpleasant / Pleasant
5. Awful / Nice

Perceived Intelligence

6. Incompetent / Competent
7. Ignorant / Knowledgeable
8. Irresponsible / Responsible
9. Unintelligent / Intelligent
10. Foolish / Sensible

국문 초록

대화형 에이전트의 심리적이고 감성적인 능력이 인간과 컴퓨터의 자연스러운 관계 형성을 위해 필요로 된다. 대화형 에이전트의 부자연스러운 표현과 반응은 사용자들에게 오히려 반감을 줄 수 있으며, 관계에 부정적인 영향을 끼친다. 감성 컴퓨팅 분야에서 주로 감정을 적용해 이를 해결했다면, 본 연구에서는 성격을 부여함으로써 대화형 에이전트의 자연스러운 피드백과 반응을 표현하고자 한다.

본 연구에서는 대화형 에이전트의 성격을 어떻게 표현할 수 있을지에 대해 탐구했다. 성격 표현 요소들로 선정된 요소들은 시각적 피드백과 언어적 요소들이다. 피험자 간 설계 방식으로, 실험을 실시했는데, 스터디 1에서는 다른 시각적 피드백들에 따른 다섯 가지 성격의 인식을 측정했다. 스터디 2에서는 다른 성별의 목소리와 언어적 요소들에 따른 다섯 가지 성격 인식을 측정했다. 또한, 특정 성격들이 업무수행에 더 적합하다는 관점을 적용하여, 스터디 3에서는 대화형 에이전트가 수행하는 과제들과 성격들에 따라 사용자들의 선호도와 인지한 지적 능력을 측정했다.

스터디 1, 2의 연구 결과에 따르면 시각적 피드백의 색깔에 상관없이 움직임 정도에 따라 사용자들이 인식하는 성격이 달라짐

을 확인할 수 있었다. 5가지 성격들 중에, 우호성(agreeableness)을 제외한 성격들에 따른 적합한 언어적 요소들을 확인할 수 있었다. 스터디 3의 연구 결과에 따르면, 대화형 에이전트가 사회적 수행 과제를 제외한 다른 과제들을 수행할 때, 창의성(openness)이 가장 선호되고, 가장 지적으로 여겨졌다. 사회적 과제를 수행하는 대화형 에이전트일 경우에만 외향성이 가장 선호되고, 지능적으로 여겨졌다.

연구 결과들에 따르면, 빠르고, 활발한 움직임의 표현 요소들이 더 뚜렷하며, 긍정적인 성격으로 인식된다. 그리고 대화형 에이전트의 성격에 대한 인식이 목소리의 성별에 따라 달라졌다. 또한, 다양하고, 표현적인 요소들을 사용하는 것이 긍정적인 성격들을 표현하기에 적합하다. 사람들이 대화형 에이전트를 인식할 때 사람들을 인식할 때와 비슷한 패턴들을 적용함을 알 수 있었다.

주요어: 성격 표현, 대화형 에이전트, 시각적 피드백, 언어적 피드백, 수행 과제

학 번: 2017-25742