Asia Pacific Journal of Information Systems
Vol. 34 No. 2 (June 2024), 585-619

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2024.34.2.585

Which Agent is More Captivating for Winning the Users' Hearts?: Focusing on Paralanguage Voice and Human-like Face Agent

SeoYoung Lee*

Professor, DongYang University, NorthSeoul Campus, Korea

ABSTRACT

This paper delves into the comparative analysis of human interactions with AI agents based on the presence or absence of a facial representation, combined with the presence or absence of paralanguage voice elements. The "CASA (Computer-Are-Social-Actors)" paradigm posits that people perceive computers as social actors, not tools, unconsciously applying human norms and behaviors to computers. Paralanguages are speech voice elements such as pitch, tone, stress, pause, duration, speed that help to convey what a speaker is trying to communicate. The focus is on understanding how these elements collectively contribute to the generation of flow, intimacy, trust, and interactional enjoyment within the user experience. Subsequently, this study uses PLS analysis to explore the connections among all variables within the research framework. This paper has academic and practical implications.

Keywords: AI Assistants, Human-like Face, Paralanguage Voice Cues, Conversational Agents, User Satisfaction, Intention to Use, Voice Interaction, CASA (Computer-Are-Social-Actors) Paradigm, Interactional Enjoyment, Flow, Intimacy, Trust

Ⅰ. Introduction

In the rapidly expanding but intensely competitive market of AI agent products, companies are vying to capture consumers' attention and establish their proprietary technologies as industry standards. The increasing preference for voice interaction among consumers, herald as the next significant platform, underscores the pivotal role of intuitive interfaces aligned with human communication patterns (Hoy, 2018).

Paralanguage voices, incorporating elements such as prosody, pitch, pause, stress, and intonation, represent a crucial component in making AI-driven con-

*Corresponding Author. E-mail: leeseoyoungann1004@gmail.com
"This study was supported by a grant from Dong Yang University in 2021 & This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2020S1A5A8041958)"

versations more natural and emotionally resonant (Orr and Sanchez, 2018). However, despite exponential growth in the AI agent industry, challenges persist in delivering an optimal user experience, with users often facing difficulties in establishing meaningful connections with AI agents.

This study focuses on two key elements – the incorporation of human-like faces and paralanguage voices into AI agents – and their significance in enhancing user satisfaction and intention to use. The rationale behind utilizing human-like faces aligns with the CASA paradigm, where the infusion of anthropomorphic features aims to make interactions more social and enjoyable. Simultaneously, the integration of paralanguage voices seeks to bridge the gap between human communication and AI interactions, making conversations more natural and expressive. Challenges in establishing meaningful connections with AI agents, such as context comprehension, monotony, and the ability to develop intimacy, underscore the need for a deeper understanding of the impact of human-like faces and paralanguage voices on user satisfaction and intention to use.

While previous studies have compared interactions with AI agents having cartoon faces to those with human faces (Johnson et al., 2000), none have explored the combined impact of face presence or absence and the use or non-use of paralanguage. This study distinguishes itself as the first to investigate these factors in a 2X2 combination, examining user reactions, flow, trust, intimacy, and interactional enjoyment for each combination. The overarching aim is to identify the most satisfying attributes in different situations, offering valuable insights for the development of conversational agents aligning with the CASA paradigm (Lee and Nass, 2010).

The core problem addressed in this study revolves around refining the user experience in AI interactions, specifically focusing on the nuanced effects of human-like faces and paralanguage voices. Despite advancements, user experience with AI agents remains suboptimal, necessitating an investigation into the nuanced effects of these features. By exploring how speech with or without paralanguage cues and the presence or absence of a human-like face influence user satisfaction and intention to use, this research aims to provide nuanced insights for the development of AI agents that users willingly engage with daily.

Research Questions

Accordingly, this research seeks to answer the following questions:

RQ1. How does interacting with an agent that has a face or without a face AI Agent affect user satisfaction and intention to use?

RQ2. How do speech with paralanguage cues and speech without paralanguage AI Agent cues affect user satisfaction and intention to use?

RQ3. How do the flow, intimacy, trust, and interactional enjoyment levels between a user and a CA affect user satisfaction and intention to use?

This study comprises two parts. Study 1 employs a two-way ANOVA to evaluate the influence of independent variables – inclusion or exclusion of paralanguage voice cues and human-like faces or without a face – on user satisfaction and intent to use. Subsequently, Study 2 utilizes partial least squares regression (PLS) to examine the interrelations among all variables within the research framework.

Ⅱ. Literature Review

2.1. CASA Paradigm

Recent advancements in software engineering have led to the creation of chatbots - artificial intelligence software capable of learning human language and emotions and interacting with natural conversations (Mhatre et al., 2016). Human-computer interactions (HCI) have become commonplace, and technology has advanced to the extent that artificial intelligence is able to interact in an independent manner. Conversational and relational computer agents are appearing capable of mimicking emotional expressions, social conversations, and other forms of human relationship behavior (Bickmore et al., 2010).

Reeves and Nass (1996) contended that the exchanges between humans and diverse media like computers, television, and new media inherently possess social and natural qualities. When a medium such as a computer presents social cues such as language use, humans perceive the medium as a social actor rather than a simple tool and apply the same social norms used in interacting with people to the medium. This happens because the human brain reacts to social cues before taking a moment to think if the cues are real or not (Nass and Moon, 2000). Recognizing computers as social actors and responding to them as a person occurs when computers have human-like characteristics, such as language use, voice, and anthropomorphized characters (Reeves and Nass, 1996).

According to the CASA paradigm, since human beings have a habit of treating media "socially" or "naturally" as if media were humans, they apply diverse social norms in their response to computer agents, which are not humans (Reeves et al., 1996). If the argument of the CASA paradigm is valid, various emotional responses in human communication processes may also appear in the interaction between humans and robots. Previous studies in the field of HRI (Human-Robot Interaction) report that people tend to show basic emotions such as trust or bonds toward artificial objects (Breazeal, 2003).

AI speakers are optimized for interactions with humans. Therefore, anthropomorphism is a critical design factor in designing AI speakers (Adam et al, 2021). This is because the human form is the best condition for interaction with humans. The CASA paradigm also emphasizes that interactions with television, new media, and computers are intrinsically as social as human communication (Reeves and Nass, 1996). Simply saying, people do not see them as mere machines but regard them as if they are human. We feel some emotions when conversing with the AI speaker. In such a case, it is crucial to evaluate multidimensionally how anthropomorphic non-human actors are similar to humans (Waytz et al., 2010). That is, the AI conversational agent's voice pitch, speed, volume, accent, pose, duration, length, language expression, and the naturalness of the conversation befitting the situation is comprehensively evaluated so that the agent is recognized as a human-like entity (Adam et al., 2021). By enhancing human-like anthropomorphism, the social response will be better. In this case, factors such as the gender of the voice, the use of natural language, or a human-like face may affect anthropomorphism.

<Table 1> provides a comprehensive summary of previous research studies that are relevant to the domain of human-machine interaction. These scholarly works span a broad spectrum, covering theoretical applications, methodological approaches, and seminal findings. Gambino et al. (2020) delved into the importance of non-verbal cues and human facial expressions in communication, underscoring the effectiveness of the CASA (Computers Are Social Actors) framework in analyzing interactions between

<Table 1> Summary of Human-like Anthropomorphic AI Agent Studies

Authors	Sample	Country	Applied Theory / Purpose of study	Variable Names Used	Methodology	Findings, Results
Gambino et al. (2020)			Paralanguage and human face			CASA has been a productive framework for studying human-machine communication, human-computer interaction, human-robot interaction, and human agent interaction.
Mhatre et al. (2016)		India	Chatbots capable of natural conversation	Proposed System, System Architecture	Experiment	This paper describes an approach to the idea of implementing web-based artificially intelligent chat-bot as a personal assistant of the user, which stimulates setting and initiating meetings of user with his clients.
Waytz et al. (2010)	5 studies	USA	It is crucial to evaluate multidimensionally how anthropomorphic non-human actors are similar to humans	Measuring Stable Behavioral Tendencies: The IDAQ (the Individual Differences in Anthropomorphism Questionnaire, or IDAQ)	Survey	This research demonstrates that individual differences in anthropomorphism predict the degree of moral care and concern afforded to an agent, the amount of responsibility and trust placed on an agent, and the extent to which an agent serves as a source of social influence on the self.
Ene et al. (2019)			User comprehensively recognizes the AI agent's voice pitch, language expression, and natural conversation appropriate to the context and evaluates it in a multidimensional manner		Experiment	The present work aims to explore the way in which the various forms of artificial intelligence affect consumer behaviour and even change our lives.
Zanbaka et al. (2006)	138	USA	Female agents tend to be more convincing than their male counterparts	Virtual human and character conditions	one of six experimental conditions in which speaker gender (male/ female) and speaker realism (human, virtual human and virtual character) are combined.	These results align with the findings of previous research that has also shown that people tend to respond to virtual humans similarly to the way they respond to real people

<Table 1> Summary of Human-like Anthropomorphic AI Agent Studies (Cont.)

Authors	Sample	Country	Applied Theory / Purpose of study	Variable Names Used	Methodology	Findings, Results
Picard (2008)	100	USA	Emotional intelligence pertains to the capacity to identify other people's emotions and react suitably to those emotions	Machines, human emotion, emotional intelligence.	Experiment	The skills given to the relational agent may bode well for maintaining longer term interactions than when the agent was lacking relational skills
Picard (2003)		USA	More effective interactions with the user can be promoted through the emotional expressions of the computer		Describing and discussion	This article raises and responds to several criticisms of affective computing, articulating state-of-the art research challenges, especially with respect to affect in human computer interaction
Kahai and Cooper (2003)	94	USA	Media richness theory	Cue multiplicity and feedback immediacy, social perceptions, message clarity, evaluation of others, task-oriented communication, decision quality	Partial Least Squares Analysis	Media rich communication improves socio emotional communication improves task quality and decision making.
Jiang and Benbasat (2007)	120	USA	Vividness and interactivity	Trust in Agent (Competence, Benevolence, & Integrity), PU, Intention to Adopt, PEOU	Partial Least Squares (PLS)	Consumers treat online recommendation agents as "social actors" and is an integral factor influencing their adoption
Csikszent mihalyi (1999)		USA	Flow theory, immersion is when the mind is completely focused in one place		Discussion	"Optimal experience" to describe those occasions where we feel a sense of exhilaration, a deep sense of enjoyment, which we cherish for long and that becomes a landmark in our lives.
Peters et al. (2018)		Sydney, NSW, Australia	Users will use technology if interacting with it satisfies their psychological need	User Experience of wellbeing – Spheres of Experience within which technology can influence wellbeing.	A model for motivation, engagement, and thriving in the user experience (metux)	a model for bridging SDT theory to technology design practice which we refer to as METUX (Motivation, Engagement & Thriving in User Experience).

humans and machines. In a similar vein, Mhatre et al. (2016) introduced a model for creating web-based AI chatbots, highlighting the significance of chatbots in enhancing user engagement. Furthermore, Waytz et al. (2010) examined how people anthropomorphize non-human entities, uncovering the degree of moral consideration, accountability, and trust placed in these entities. These earlier studies shed light on the complexities of human-machine communication, laying the groundwork for the current research's methodology and theoretical underpinnings in exploring this field. The CASA framework, rooted in the media equation doctrine, articulates the dynamic interaction between humans and technology, treating machines as social counterparts. Despite facing criticism, the framework awaits an expansion to mirror the evolving landscape of human behaviour, technological advancement, and the intricate dance of human-technology interaction. Gambino et al. (2020), advocate for a nuanced understanding that transcends the mere transposition of social behaviours from human-human interactions to technology. They propose that individuals might be crafting distinct social conventions for engaging with digital and robotic agents. This refined viewpoint not only clarifies previously incongruent findings but also deepens our grasp of human-machine interaction across diverse fields such as artificial intelligence and digital communication.

Cacioppo and Epley (2010)'s work presents a thorough investigation across five studies in the United States, focusing on the multidimensional nature of anthropomorphism. The introduction of the Individual Differences in Anthropomorphism Questionnaire (IDAQ) illuminates how personal propensities towards anthropomorphism significantly dictate the moral value, trust, and accountability ascribed to non-human actors. This revelation underscores the profound impact these anthropomorphized agents have on individuals' social conduct and decision-making processes, suggesting that the extent of human-like attributes assigned to machines significantly influences their interaction with and perception of these entities. Ene et al. (2019)'s research employs a comprehensive qualitative and quantitative lens to scrutinize consumer perceptions of anthropomorphic AI designs. This meticulous examination reveals how users assess AI agents, taking into account voice pitch, linguistic expression, and the capacity for contextually relevant conversations. By adopting experimental methodologies, this study endeavours to demystify the effect of varied AI forms on consumer behaviour and the overarching changes they herald in our daily existence. This holistic approach unravels the subtle ways in which anthropomorphic AI designs sway user experience, engagement, and acceptance, offering pivotal insights for tailoring AI to seamlessly align with consumer desires and expectations.

2.2. Anthropomorphism (Anthropomorphic Properties)

When AI agents employ anthropomorphic properties such as a human-like voice or language expressions and empathetic dialogue methods, users' psychological resistance to the system is relatively low (Dolganov and Letnev, 2020). This is because, in communicating with the AI agent, the user comprehensively recognizes the AI agent's voice pitch, language expression, and natural conversation appropriate to the context and evaluates it in a multidimensional manner (Ene et al., 2019).

Emotions are essential conditions for the formation and maintenance of human relationships. They increase favourable feelings for others through the communication process, form a sense of companionship, and provide social support, which is a psychological resource obtained through relationships (Coursaris and Liu, 2009; Ginossar, 2008). Therefore, an AI agent that expresses emotions can make users perceive the AI agent to be a relational object by forming empathy and a sense of companionship and providing social support. Coeckelbergh (2011) emphasized that a robot, such as an AI agent, needs companionship to become a member of us, and empathy is important in forming such companionship. In addition, humans empathize with other people because they have similar lives and problems and have a similarity of vulnerability, and this similarity is necessary for AI agents to empathize with humans. Therefore, if the AI agent can express similar emotions to a user, it will attract the user's attention more and positively affect the use of AI for learning, information retrieval, and entertainment. As such, emotional intelligence capable of identifying and expressing the user's emotional state is required (Hou and Lee, 2011). In this context, emotional intelligence pertains to the capacity to identify other people's emotions and react suitably to those emotions (Picard, 2007). With the development of AI, in the near future, AI agents will be able to understand users' emotions and respond appropriately. Therefore, the emotional intelligence of computers is expected to become increasingly important (Minsky, 2007). According to Zanbaka et al. (2006), female agents tend to be more convincing than their male counterparts.

2.2.1. Conversational Agents with Faces/without Face

Nonverbal communication, particularly through facial expressions, plays a pivotal role in human-human interactions, conveying emotions, intentions, and establishing the tone of communication (Bartel and Savedra, 2000). The integration of faces into conversational agents in the field of artificial intelligence (AI) has become an area of increasing interest, aiming to replicate and enhance the nonverbal cues that humans naturally employ in communication. This literature review explores the significance of nonverbal cues in conversational agents, focusing on facial expressions and their impact on user perception.

Mehrabian (1971) emphasize the potency of facial expressions in conveying messages related to attractiveness, strength, and honesty. The orientation of a face, whether sideway or facing the listener, can communicate distinct impressions. A sideway-oriented face may convey shame or ignorance, while a face directly engaging the listener imparts interest and concentration. It is revealed that employing a combination of cues is more effective than relying on a single cue, as mimicking single cues may elicit revulsion, a phenomenon known as the uncanny valley in robotic research (Mori, 1970). As technology evolves, it is essential to address the unexplored territory of mobile environments and animated agents, ensuring the applicability and effectiveness of these communication tools across diverse platforms.

Given the research results indicating that interactions are more favourable when AI agents exhibit human-like features according to the CASA paradigm, I hypothesize that the presence or absence of agents with human-like faces will influence user's flow, trust, intimacy, and interactional enjoyment. This assumption is grounded in the belief that the more an AI agent resembles a human, the more positively users will perceive and engage in interactions, aligning with the principles of the CASA paradigm.

H1: AI agents with a face will have a positive (+) effect on the user's flow.
H2: AI agents with a face will have a positive (+) effect on the user's trust.
H3: AI agents with a face will have a positive (+) effect on the user's intimacy.
H4: AI agents with a face will have a positive (+) effect on the user's interactional enjoyment.

2.2.2. Voice (Delivery Power)

In HCI design, the emotional aspect of the computer is very important because more effective interactions with the user can be promoted through the emotional expressions of the computer (Picard, 2003). However, what is important here is not simply whether the computer agent expresses emotions or the extent to which emotions are expressed but having the ability to understand the user's emotional state and express timely emotions that fit the state (Hou and Lee, 2011). Voice is the most effective way to express emotions. Since interactions through voice are unique and natural to humans, it doesn't require training to use this modality to interact with devices and is more efficient than existing methods of inputting texts by hand (Back et al., 2012).

People anthropomorphize non-human actors such as social robots by giving them appearances, acts, and characteristics similar to those of humans, including not only various facial expressions but also genders and voices that fit the identity (Chandler and Schwarz, 2010). For example, people make certain preconceived judgments about computers based on a computer-generated voice. A computer employing a male voice is regarded as more logical compared to one with a female voice. Conversely, a computer featuring a female voice is deemed to be comparatively more emotional (Nass and Moon, 2005). According to a survey on AI robots, although 56% of gendered AI robots are female, the majority of "bots" used in the fields of "law" or "finance" are male (Kelshaw, 2016).

2.2.3. Media Richness

Although face-to-face communication is the smoothest mode of communication, people use media to communicate when it is inconvenient to gather people together (Hessels et al., 2019). Communication through media differs greatly depending on the characteristics of the media. According to the media richness theory, media that facilitate seamless communication, such as face-to-face settings, are labeled as 'rich' media, while other forms of media are referred to as 'lean' media (Lee and Borah, 2020).

Research on media richness indicates that video, being the closest to face-to-face communication, ranks as the richest medium, followed by audio, with texting being lean (Kahai and Cooper, 2003). Communication is more accurate and richer where the medium provides immediate feedback or can transmit vivid messages able to stimulate various sensory organs (Daft and Lengel, 1986; Kahai and Cooper, 2003; Jiang and Benbasat, 2007).

Although various elements are known as characteristics that increase media richness, vividness and interactivity are particularly prominent characteristics (Zhu et al., 2010). Vividness means the degree to which a medium can deliver messages and clues through diverse sensory organs, and interactivity refers to the degree to which the medium user can manipulate and receive messages transmitted by the medium (Jiang and Benbasat, 2007).

In face-to-face communication, meanings can be more accurately understood through various clues such as facial expressions, gestures, and nuances. Media richness can elicit emotional responses from media users, such as a more positive experience of using media (Kahai and Cooper, 2003). Therefore, face-to-face communication can elicit higher cognitive responses, such as focused attention and recalled knowledge, and positive emotional responses, such as flow into the media use situations and perceived enjoyment (Suh et al., 2005).

2.2.4. Paralanguage

Among the elements of message delivery, the content of the message accounts for 7%, while the voice (tone, intonation, and volume) accounts for about 38% (Mehrabian, 1969). In message delivery, the voice that delivers the message is more important than the content. Paralanguage is performed through the voice among the actions of nonverbal carriers involved in communication. It refers to elements related to various forms of sound except for formal language in the communication process. In other words, paralanguage refers to phonetic and phonological elements, excluding pure language, as they are expressed in communication.

In addition, paralanguage can be defined as "elements associated with various forms of sound" that perform functions to increase the communication effectiveness of a language (Tepper and Haase, 1978). Leigh and Summers (2013) view paralanguage as a type of nonverbal communication and emphasize the effectiveness of communication through paralanguages.

In other words, they suggest that speech hesitation negatively affects the speaker's trustworthiness, ability, professionalism, persuasion, and interest Littlejohn and Foss (2010) argue that paralanguage expresses the speaker's personality and emotions in the communication process, profoundly affecting the listener's level of understanding and persuasion. Paralanguage factors positively affect the perception of the speaker's ability and sociability, increasing trust. On the other hand, tone, speed, and volume, which are types of paralanguages, greatly influence the delivery of messages. In general, a loud voice gives the impression of overwhelming the listener, and a gentle voice gives the impression of submissiveness, which has a negative effect on trust. In addition, harsh tones can irritate the listener's psychology, causing rejection, and an excessively weak voice can lead to doubt. On the other hand, moderately low tones can soothe the listener's mind and increase trust (Wainwright, 1999). Most paralanguage reinforces the linguistic message and verbal expression but also communicates itself (Dash and Davis, 2022). These paralanguages replace or facilitate verbal communication and express our emotions and positions. Therefore, such paralanguage should be well expressed for smooth interaction between AI agents and users. The components of paralanguage are pronunciation, intonation, speed, and pause (Dash and Davis, 2022).

Pronunciation is an important factor in delivering a message and involves the use of speech organs such as the tongue, lips, and teeth to produce speech sounds. It plays an essential role in correctly expressing the speaker's thoughts or feelings.

Intonation refers to the pitch of a note. It is the melodic pattern of an utterance in phonetics and conveys differences in expressive meaning, such as surprise, anger, or happiness. It involves variation in the pitch of a voice. For instance, rhythm and stress are often accompanied to convey a particular meaning in English.

Speed refers to the rate of speech, which varies according to the content and emotion of the sentence. The speed of speech reflects emotions and attitudes. People who speak quickly convey excitement and are expressive and persuasive. However, speaking too quickly makes the listener nervous and insecure. Therefore, it is important to communicate with an appropriate speed of speech for the situation in communication.

Pause means to stop speaking momentarily. In other words, it is difficult to have a conversation in one breath because the sentences are long, or a pause is put for a while due to problems with meaning or pronunciation. However, it is essential to say the utterance properly to convey its meaning.

2.2.5. Flow

Flow can be conceptualized as a situation in which a person who has a goal to achieve fully concentrates on something, even with their mindset (Davis and Wong, 2007). This concept has seen extensive use across diverse domains like virtual reality, psychology, and online gaming (Csikszentmihalyi, 1975). Users of AI agents experience flow, which is a cognitive state where they talk with an AI agent as if it is a human without feeling any sense of distance from the AI agent while using the service. This cognitive flow results from AI agent interactions (Agarwal and Karahanna, 2000). In addition, flow can affect the user's attitude and behavior, resulting in a positive relationship that positively affects trust and intimacy with AI agents (Ho and Kuo, 2010).

From a business psychology viewpoint, emotional engagement and the degree of flow determine consumer product preferences. According to flow theory, immersion is when the mind is completely focused in one place (Csikszentmihalyi, 1991). Novak's research on Internet consumer immersion sees flow as a factor directly affecting purchase (Novak et al., 2000). Flow theory makes sense when considering the features of speech recognition technology. Elements that create emotional flow may be induced by chit-chatting or by paralanguage. Therefore, AI agent service providers must raise the level of personification of AI agents so that users can enjoy and flow into their service.

'Flow' is a concept defined by Csikszentmihalyi (1999), and it expresses the feeling that actions are carried out naturally as water flows at the moment when life is heightened. In other words, it means a state full of conscious experiences, such as feeling comfortable like water flowing when immersed.

H5: AI agents with paralanguage voice will have a positive (+) effect on the user's flow.
H9: The user's flow to the AI agent will have a positive (+) effect on user satisfaction.

2.3. Trust

Trust has been explored in numerous scholarly fields, including psychology, sociology, and economics. It pertains to the conviction that an individual or technology possesses the required qualities to accomplish the anticipated outcomes (Mayer et al., 1995). In particular, trust in technology is established when, like trust in humans, there is an expectation that the technology's ability is premised on objective characteristics. For example, users' trust in AI agents can be formed when AI agents continuously operate, although they have no will, or when they show personified attributes like the attributes of humans. Therefore, the scope of application of trust has been expanded to include not only interactions between humans and society, such as individuals, groups, and organizations, but also human-computer interactions (HCI), such as the relationships among the members of virtual teams and e-commerce (Song et al., 2020). As technologies such as information technology and artificial intelligence have increased in human life, discussions on trust have extended beyond trust in humans to trust in technology. Studies on trust in technologies conducted in the 2000s showed that trust in various technical artifacts, such as AI recommendation agents, had remarkable effects on user behavior (Wang and Benbasat, 2005).

H6: AI agents with paralanguage voice will have a positive (+) effect on the user's trust.
H10: The user's trust in the AI agent will have a positive (+) effect on user satisfaction.

2.4. Intimacy

Intimacy is one of the predisposing factors for forming trust. People tend to trust close friends more and are more easily persuaded by them (Rempel et al., 1985). Charles Berger's uncertainty reduction theory explains that uncertainty about others decreases in the process of becoming intimate. In addition, intimacy includes feeling emotionally close and behaviours, as well as various psychological and cognitive changes that appear in intimate relationships (Aron et al., 1992). Morkes et al. (1998) found that users feel higher levels of goodwill towards computers that use humour and evaluate such computers as capable and cooperative, and Bickmore (2001) argued that awareness of computers' interactions, such as chats, could be a factor that makes users trust the technology more. Therefore, it can be seen that the personification of AI agents can enhance user intimacy and trust.

H7: AI agents with paralanguage voice will have a positive (+) effect on the user's intimacy.
H11: The user's intimacy with the AI agent will have a positive (+) effect on user satisfaction.

2.5. Interactional Enjoyment

The core characteristic of an AI agent service is smooth communication between AI agents and service users. Such communication can be defined as interactivity; through this interaction, users can be satisfied with and enjoy the service (Adam et al., 2021). Feeling the AI agent as a human being and interacting with the anthropomorphic AI agent is one major factor that increases the user's enjoyment and continuous use of the AI agent (Park et al., 2018). In addition, when visual anthropomorphism is attempted, such as a smiling face, even in the case of an error situation, higher levels of enjoyment and intention to use can be shown compared to the situation without anthropomorphism (Qiu and Benbasat, 2009).

H8. AI agents with paralanguage voice will have a positive (+) effect on the user's interactional enjoyment.
H12. The user's interactional enjoyment of the AI agent will have a positive (+) effect on user satisfaction.

2.6. User Satisfaction

Satisfaction is defined as the subjective evaluation of results or experiences obtained through the process of consuming products (Maxham, 2001). In essence, satisfaction pertains to the psychological and emotional state stemming from the disparity between a consumer's product expectations and their assessment of the outcomes. Previous research has found that when a user has a positive experience with a particular product or system, it leads to the formation of user satisfaction, which in turn has a crucial impact on the intention to persist in using the product or service.

Therefore, user satisfaction is an important leading variable in identifying the intention to continue using (Carroll and Ahuvia, 2006). Rafaeli (1988) revealed that satisfaction is potentially related to interactivity and said that the higher the interactivity of a certain product, the higher the user satisfaction with the product. Sundar (2005) argued that as interactivity increases, user satisfaction with the function of the relevant product increases.

2.7. Intention to Use

Intention to use refers to the user's intention to continue using a certain system after trying it for the first time. Users' continuous use of new products is closely related to securing the competitive advantage of the products in the market, along with the success of the product system. In particular, since AI agents are big data-based artificial intelligence, users need to interact continuously with the AI agents. According to Peters et al. (2018), the self-determination theory predicts that users will use technology if interacting with it satisfies their psychological needs. Concurrently, satisfaction with use and the intention to use are closely connected, with numerous prior studies demonstrating that satisfaction with usage positively influences the intention to continue use. Davis (1989) asserted that the intention to employ new technology could foresee its actual utilization. He further mentioned that the intention to adopt new technology is influenced by one's attitude towards it and clarified that perceived usefulness and ease of use contribute to shaping this attitude.

H13. User satisfaction with the AI agent will have a positive (+) effect on the intention to use.

Ⅲ. Method

3.1. Participants and Procedure

An experimental study was conducted based on the type of AI agent. A total of 640 (160 × 4) adult men and women aged from their 20s to over 50s, living in Seoul and other cities, participated in the experiment. The participants completed a survey for their respective experiments in their AI agent group. <Table 2> lists the respondents' demographic characteristics. The questionnaires used a 7-level Likert scale ranging from "(1) completely disagree" to "(7) completely agree." An experimental environment was created to collect information from the participants concerning each AI agent group. Four types of AI agents were set, and the participants were randomly assigned to one of the four groups. Participants randomly assigned to one of the four conditions in a 2 (AI agent with face vs No Face) x 2 (AI agent with paralanguage voice vs a computer voice) between-subjects design. The four types are as follows:

3.2. Research Model and Hypotheses

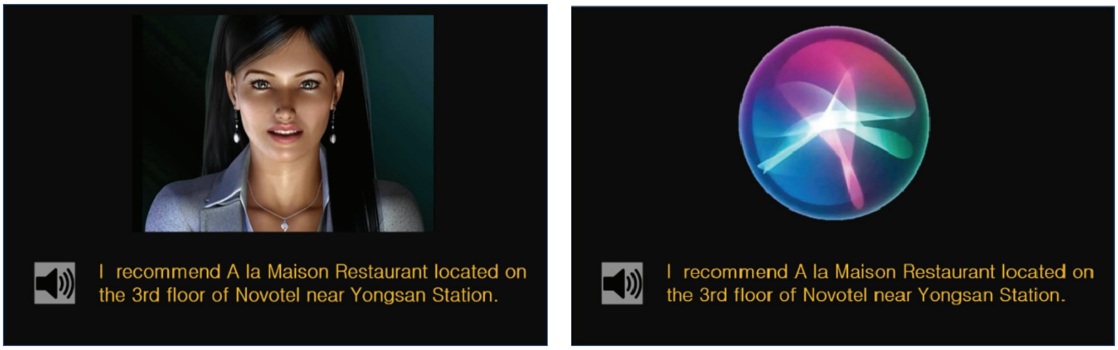
Users were more engaged when interacting with a talking face than with text-based interfaces and spent more time with the talking face agent (Walker et al., 1994). Interfaces with a face were more satisfying and natural (Sproull et al., 1996).

Female agents were selected over male agents in

<Table 2> Demographic Information of Respondents

Category (Group)		A	B	C	D
Sex	Male	78	81	75	77
	Female	82	79	85	83
Age	19-29	70	69	65	58
	30-40	68	68	64	60
	40-50	12	13	20	30
	50+	10	10	11	12
City	Seoul	55	56	53	54
	Pusan	43	35	40	37
	Daegu	33	36	29	28
	Daejeon	14	22	25	21
	Gwangju	15	11	13	20

Note: A Type 1: AI agent with face + AI agent with paralanguage voice.
B Type 2: AI agent with no face + AI agent with paralanguage voice.
C Type 3: AI agent with face + AI agent with a computer voice.
D Type 4: AI agent with no face + AI agent with computer voice.



<Figure 1> Example of AI Agent Types (AI Agent with Face vs. AI Agent with No Face and Computer Voice)

this study, as prior research indicated that they tend to be more persuasive (Zanbaka et al., 2006). Attractive agents are seen as more likeable and convincing. Users found anthropomorphic agents more attractive and trustworthy, with feminine avatars being perceived as more appealing than their masculine counterparts. Similarly, users regarded anthropomorphic human agents as more attractive and credible, with female agents seen as more alluring than male agents (Nowak and Rauh, 2005; Nowak and Rauh, 2008). In this research, I adopted an attractive female agent with a human paralinguistic voice.

This study consists of two parts. Study 1 involves a two-way analysis of variance (ANOVA) to determine the effects of independent variables, namely, agent with face and voice with paralanguage, on user satisfaction and intention to use. Then, Study 2 uses partial least squares regression (PLS) to investigate the relationships among all variables in the research model. These relationships are verified by testing the hypotheses derived from theoretical relevance. The rationale for conducting this study lies in the need to elucidate how the outcomes of variables may vary based on the 2x2 combinations of faces (presence or absence) and paralanguage cues (presence or absence), forming four distinct groups. By establishing these four groups, the experiment aims to reveal the impact of these combinations on the results. The experimental design, encompassing the nuanced interactions of face and paralanguage variables, enables a comprehensive exploration of their effects. This experimental approach is justified by the fact that it allows for a systematic examination of the influence of human-like faces and paralanguage voices on user experiences. The formation of four groups ensures a controlled exploration of each factor's individual and combined effects. Such experimentation provides a robust foundation for understanding the intricate dynamics between these variables, contributing to the validity and reliability of the study. In summary, the experiment is crucial for unraveling how variables' outcomes change within the context of the 2x2 combinations, shedding light on the significance of human-like faces and paralanguage voices in user interactions. This experimental design enhances the validity and relevance of the study by providing a structured and controlled exploration of these influential factors.

The study aimed to validate the structural relationship between the influence of an AI agent's facial appearance and voice type on flow, trust, intensity, user interactional enjoyment, and user satisfaction and intention to use through the development of the following research model.

H1. AI agents with a face will have a positive (+) effect on the user's flow.

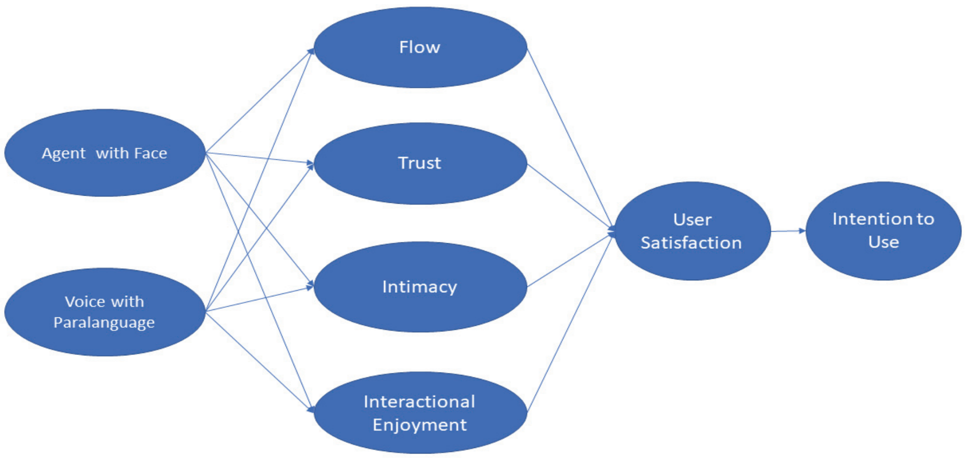
H2. AI agents with a face will have a positive (+) effect on the user's trust.

H3. AI agents with a face will have a positive (+) effect on the user's intimacy.

H4. AI agents with a face will have a positive (+) effect on the user's interactional enjoyment.

H5. AI agents with paralanguage voice will have a positive (+) effect on the user's flow.

H6. AI agents with paralanguage voice will have a



<Figure 2> Research Model

positive (+) effect on the user's trust.

H7. AI agents with paralanguage voice will have a positive (+) effect on the user's intimacy.

H8. AI agents with paralanguage voice will have a positive (+) effect on the user's interactional enjoyment.

H9. The user's flow to the AI agent will have a positive (+) effect on user satisfaction.

H10. The user's trust in the AI agent will have a positive (+) effect on user satisfaction.

H11. The user's intimacy with the AI agent will have a positive (+) effect on user satisfaction.

H12. The user's interactional enjoyment of the AI agent will have a positive (+) effect on user satisfaction.

H13. User satisfaction with the AI agent will have a positive (+) effect on the intention to use.

3.3. Data Processing

T-test was performed to verify the difference between flow, trust, intimacy, and interactional enjoy-

<Table 3> Measurement Items of Research Constructs

Construct	Survey Items	Reference
Flow	1. I was fully immersed in the interaction and lost track of time. 2. The interaction with the AI agent was so absorbing that I became fully engrossed in the experience. 3. I found the interaction with the AI agent inherently rewarding and enjoyable. 4. I found the interaction with the AI agent inherently rewarding and enjoyable.	Agarwal and Karahanna (2000); Csikszentmihalyi (1991); Csikszentmihalyi (1999); Davis and Wong (2007); Ho and Kuo (2010); Novak et al. (2000)
Intimacy	1. I feel that the CA is my close friend. 2. I feel emotionally close to the CA. 3. I developed a sense of familiarity with the CA.	Aron et al. (1992). Berschied et al. (1989)
Trust	1. I have faith in what the CA is telling me. 2. The CA provides with me unbiased and accurate movie recommendations. 3. The CA is honest. 4. The CA is trustworthy 5. I believe that the CA provides a reliable service. 6. I can trust the CA with my personal information. 7. I can trust the information provided by the CA.	Dinev and Hart (2006); Morgan and Hunt (1994); Moorman et al. (1993) Wang and Benbasat (2005)
Interactional Enjoyment	1. It is fun and enjoyable to share a conversation with the CA. 2. The conversation with the CA is exciting. 3. I enjoy more if it was recommended by the CA than when I choose it myself. 4. Services provided by the CA are more entertaining and attractive than without a CA.	Koufaris (2002); Van der Heijden (2003, 2004)
User Satisfaction	1. I was satisfied with the experience of using a dialogue with the CA to complete tasks. 2. Interacting with the CA was a pleasant and satisfactory experience. 3. The dialogue with the CA gave me useful information. 4. The overall assessment of conversing with the CA was satisfactory.	Brill et al. (2022); Chin et al. (1988)
Intention to Use	1. I will use the CA system again. 2. If this CA system is commercially available, I would purchase it.	Davis et al. (1992); Wang and Benbasat (2005)

ment according to the presence or absence of the AI agent's face (face or mike) and voice type (paralanguage or electric voice). Moreover, ANOVA analysis was conducted to confirm the differences between groups based on the presence or absence of facial features and voice type, while Scheffe's test was utilized for the post-hoc examination. To verify the measurement tool's validity and reliability, exploratory factor analysis was executed, and the convergent validity of the research model was assessed by evaluating the construct reliability (CR) and average variance extracted (AVE). Lastly, structural equation model analysis was carried out to investigate the structural relationships among the research variables. All statistical analyses were completed using a significance level of 0.05.

3.4. Measurement

Flow is characterized by a deep and effortless concentration on the task at hand. The individual is fully absorbed and immersed in the activity (Csikszentmihalyi, 1990).

Intimacy includes feeling emotionally close and behaviors, as well as various psychological and cognitive changes (Aron et al., 1992).

Trust is generally regarded as a psychological mechanism for reducing uncertainty and increasing the likelihood of a successful (e.g., safe, pleasant, satisfactory) interaction with entities (Lukyanenko et al., 2022).

Interactional enjoyment of using AI agents refers to the positive feelings of happiness and pleasure that users experience when interacting with artificial intelligence (AI) agents (Pelau et al., 2023).

User satisfaction is a measure of how well a user's expectations are met when interacting with artificial intelligence (AI) systems (Ruel and Njoku, 2021).

Intention to use technology is a degree to which the user would like to use technology in the future (Hussein, 2015).

Ⅳ. Results

4.1. Study 1.

4.1.1. Verification of differences in flow, trust, intimacy, and interactional enjoyment between the exposure and non-exposure of the face

T-tests were conducted to analyze the differences in flow, trust, intimacy, and interactional enjoyment between the exposure and non-exposure of the agent's face, and the results are shown in <Table 4>. First, when the flow of users was examined, the mean value (M) of flow was shown to be 5.270 (SD = 1.416) when the agent's face was exposed and 4.206 (SD = 1.564) when not exposed, and the difference was statistically significant ($t = 9.267$, $p < .001$). That is, it can be seen that the level of user's flow is higher when the agent's face is exposed than when not exposed. Next, when the user's trust was examined, the mean value of the trust (M) was 5.476 (SD = 1.165) when the face was exposed and 4.527 (SD = 1.536) when the face was not exposed, and the difference was shown to be statistically significant ($t = 8.439$, $p < .001$). That is, it can be seen that the level of trust of users is higher when the agent's face is exposed than when not exposed. Next, when the user's intimacy was examined, the mean value (M) of intimacy was shown to be 5.220 (SD = 1.730) when the face was exposed and 4.180 (SD = 1.785) when the face was not exposed and the difference was statistically significant ($t = 7.737$, $p < .001$). That

<Table 4> Verification of Differences in Flow, Trust, Intimacy, and Interactional Enjoyment between the Exposure and Non-exposure of the Face

DV	face	M	SD	t	p
Flow	Y	5.284	1.402	9.436	.000
	N	4.203	1.560		
Trust	Y	5.503	1.156	9.351	.000
	N	4.525	1.529		
Intimacy	Y	5.235	1.721	7.861	.000
	N	4.172	1.782		
Interactional enjoyment	Y	5.231	1.478	9.417	.000
	N	4.113	1.595		

is, it can be seen that the level of the user's intimacy is higher when the agent's face is exposed than when not exposed. Finally, when the user's interactional enjoyment was examined, the mean value (M) of interactional enjoyment was shown to be 5.216 (SD = 1.491) when the face was exposed and 4.116 (SD = 1.599) when not exposed, and the difference was statistically significant ($t = 9.244$, $p < .001$). That is, it can be seen that the level of interactional enjoyment of users is higher when the agent's face is exposed than when not exposed.

4.1.2. Verification of differences in flow, trust, intimacy, and interactional enjoyment between the exposure and non-exposure of the paralanguage voice

T-tests were conducted to analyze the differences in flow, trust, intimacy, and interactional enjoyment between the agent's voice types, and the results are shown in <Table 5>. First, when the flow of users was examined, the mean value (M) of flow was shown to be 5.473(SD = 1.162) when the voice was a pseudo-language and 4.007(SD = 1.609) when the voice was a machine sound, and the difference was shown to be statistically significant ($t = 13.560$, $p < .001$). That is, it can be seen that the level of user's flow is higher when the agent's voice is a paralanguage than when it is a machine sound. Next, when users' trust was examined, the mean value of the trust (M) was 5.476 (SD = 1.165) when the voice was a paralanguage and 4.478 (SD = 1.535) when the voice was a machine sound, and the difference was shown to be statistically significant ($t = 9.505$, $p < .001$). That is, it can be seen that the level of trust of users is higher when the agent's voice is a paralanguage than when the voice is a mechanical sound. After that, when the user's intimacy was examined, the mean value (M) of intimacy was shown to be 5.580 (SD = 1.323) when the voice was a paralanguage and 3.820 (SD = 1.852) when the voice was a machine sound, and the difference was shown to be statistically significant ($t = 14.174$, $p < .001$). That is, it can be seen that users' intimacy level is higher when the agent's voice is a paralanguage than when it is a machine sound. Finally, when the user's interactional enjoyment was examined, the mean value (M) of interactional enjoyment was shown to be 5.313 (SD = 1.352) when the voice was a paralanguage and 4.022 (SD = 1.650) when the voice was a machine sound, and the difference was shown to be statistically

<Table 5> Verification of Differences in Flow, Trust, Intimacy, and Interactional Enjoyment between the Types of Voices

DV	Voice	M	SD	t	p
Flow	Y	5.492	1.143	13.956	.000
	N	3.995	1.599		
Trust	Y	5.558	1.109	10.564	.000
	N	4.470	1.527		
Intimacy	Y	5.598	1.305	14.518	.000
	N	3.809	1.842		
Interactional Enjoyment	Y	5.334	1.337	11.456	.000
	N	4.011	1.640		

significant ($t = 11.115$, $p < .001$). That is, it can be seen that the level of interactional enjoyment of users is higher when the agent's voice is a paralanguage than when it is a machine sound.

ANOVAs were conducted to analyze the differences in flow, trust, intimacy, and interactional enjoyment among four groups divided according to whether the face was exposed or not (human face/microphone) and the types of voice (paralanguage/machine sound) and the results are as shown in <Table 6>. First, when the flow of users was examined, the mean values (M) of flow were shown to be significantly different among group A (the agent's voice was a paralanguage and the face was exposed), group B (the agent's voice was a paralanguage and the face was not exposed), group C (the voice was a machine sound and the face was exposed), group D (the agent's voice was a machine sound and the face was not exposed) ($F = 116.894$, $p < .001$). Scheffe's post hoc tests were conducted to compare the user's flow level among the individual groups, and the results showed that group A's flow level was the highest, followed by group B, group C, and group D in order of precedence. Scheffe's post hoc tests were conducted to compare the levels of user trust among the individual groups, and the results showed that group A's trust level was higher than group B's. The trust level of group C was higher than that of group D. Scheffe's post hoc tests were conducted to compare the levels of user intimacy among the individual groups, and the results showed that the level of intimacy of group A was the highest, followed by group B, group C, and group D in order of precedence.

Finally, when the user's interactional enjoyment was examined, the mean values (M) of interactional enjoyment were shown to be significantly different among group A, which is the case where the agent's voice was a paralanguage, and the human-like face was exposed, group B, which is the case where the agent's voice was a paralanguage and the face was not exposed, group C, which is the case where the voice was a computer voice, and the human-like face was exposed, Group D, which is the case where the agent's voice was a computer voice and the human-like face was not exposed ($F = 87.608$, $p < .001$). Scheffe's post hoc tests were conducted to compare the levels of user interactional enjoyment among the individual groups, and the results showed that the level of interactional enjoyment of group A was higher than that of group B, and the level of interactional enjoyment of group C was higher than that of group D.

<Table 6> Verification of Differences among Groups Divided according to Whether the Face is Exposed or not and Voice Types

DV	Group	M	SD	F	p	Post hoc (Scheffe)
Flow	A	5.825	1.225	125.093	.000	A > B > C > D
	B	5.159	0.947			
	C	4.742	1.362			
	D	3.248	1.467			
Trust	A	5.829	1.229	87.123	.000	A > B, C > D
	B	5.287	.901			
	C	5.177	.979			
	D	3.762	1.646			
Intimacy	A	5.910	1.352	113.720	.000	A > B > C > D
	B	5.285	1.179			
	C	4.559	1.787			
	D	3.059	1.577			
Interactional Enjoyment	A	5.692	1.415	93.533	.000	A > B, C > D
	B	4.976	1.152			
	C	4.770	1.389			
	D	3.251	1.510			
User Satisfaction	A	5.845	1.239	90.496	.000	A > B, C > D
	B	5.159	1.036			
	C	5.108	1.132			
	D	3.633	1.596			
Intention to Use	A	5.851	1.416	91.268	.000	A > B, C > D
	B	5.164	1.117			
	C	4.994	1.506			
	D	3.309	1.743			

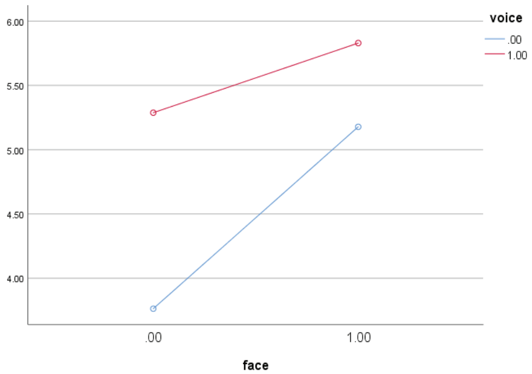
Two-way ANOVA was performed to analyze the interaction effect of the presence/absence of the human-like AI agent face and the presence/absence of paralanguage voice. <Table 7> shows the results of analyzing the interaction effect between the presence of the AI agent face and the voice type on the user's trust. It can be seen from <Table 7> that the difference in user's trust according to the presence or the absence of a face on trust is examined as (F = 107.294, $p < .001$), and the presence/absence of paralanguage voice on trust is examined as (F = 132.703, $p < .001$). From these tables, it can be seen that all these parameters significantly affect the user's trust. The interaction effect between the presence of the agent's human-like face and the paralanguage voice type on trust was examined (F = 21.374, $p < .001$) to be significant. <Figure 3> shows a graphical illustration of the interaction effect between the presence or absence of a face and the voice type, and it can be seen that the user's trust in the AI agent improved when the AI agent has a face and a paralanguage voice.

<Table 7> Two-way ANOVA (Trust)

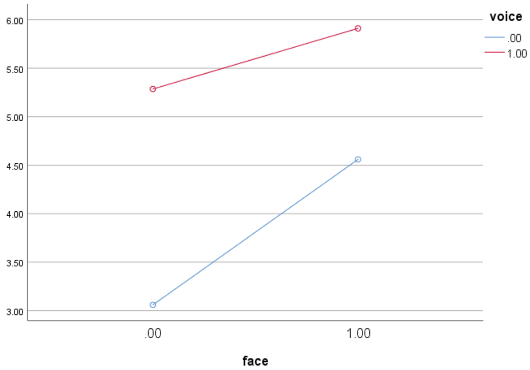
	SS	df	MS	F	p
Face	160.793	1	160.793	107.294	.000
Voice	198.873	1	198.873	132.703	.000
Face*voice	32.031	1	32.031	21.374	.000

<Table 8> Two-way ANOVA (Intimacy)

	SS	df	MS	F	p
Face	189.656	1	189.656	85.213	.000
Voice	537.501	1	537.501	241.500	.000
Face*voice	32.156	1	32.156	14.448	.000



<Figure 3> Interaction Effects of Face on Trust



<Figure 4> Interaction Effects of Face on Intimacy

The results of analyzing the interaction effect between the presence of the AI agent's face and the voice type on the user's intimacy are shown in <Table 8>. As shown in <Table 8>, the difference in user's intimacy according to the presence or absence of a face ($F = 85.213$, $p < .001$) and the type of voice ($F = 241.500$, $p < .001$) is examined. It can be seen that all significantly affect the user's intimacy, and the interaction effect between the presence of the face and the voice type ($F = 14.448$, $p < .001$) was also significant. <Figure 4> is a graph showing the interaction effect between the presence or absence of a face and the voice type, and it can be seen that the user's intimacy with the AI agent improved when the AI agent had a face and a paralanguage voice.

The results of analyzing the interaction effect between the presence of the AI agent's face and the voice type on the user's flow are shown in <Table 9>. As shown in <Table 9>, the difference in user's flow according to the presence or absence of a face ($F = 122.371$, $p < 0.001$) and the type of voice ($F = 234.964$, $p < 0.001$) is examined. It can be seen that all have a si significant effect on the user's flow, and the interaction effect between the presence of the face and the voice type ($F = 17.943$, $p < .001$) was also significant. <Figure 5> shows the interaction effect between the presence or absence of a face and the voice type, and it can be seen that the user's flow in the AI agent improved when the AI agent has a face and a paralanguage voice.

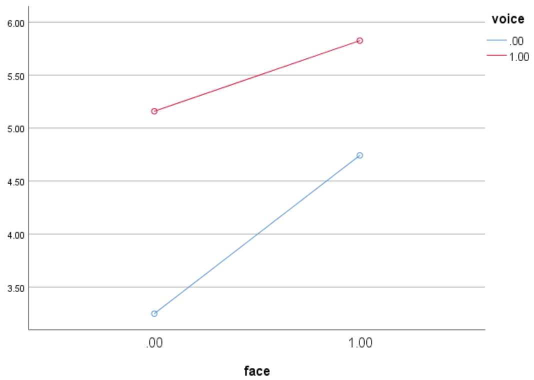
The results of analyzing the interaction effect between the presence of the AI agent's face and the voice type on the user's interactional enjoyment are shown in <Table 10>. As shown in <Table 10>, the difference in users' interactional enjoyment according to the presence or absence of a face ($F = 110.885$, $p < .001$) and the type of voice ($F = 155.381$, $p < .001$) is examined. It can be seen that all significantly affect the user's interactional enjoyment, and the in-

<Table 9> Two-way ANOVA (Flow)

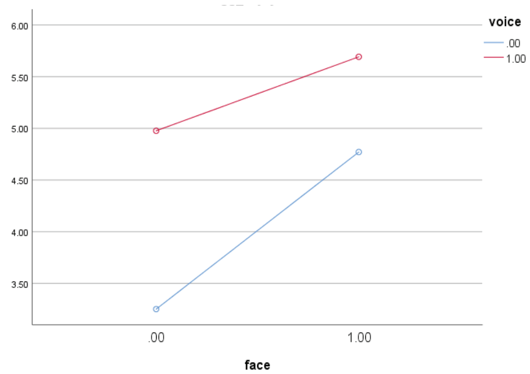
	SS	df	MS	F	p
Face	196.085	1	196.085	122.371	.000
Voice	376.501	1	376.501	234.964	.000
Face*voice	28.751	1	28.751	17.943	.000

<Table 10> Two-way ANOVA (Interactional Enjoyment)

	SS	df	MS	F	p
Face	209.822	1	209.822	110.885	.000
Voice	294.018	1	294.018	155.381	.000
Face*voice	27.121	1	27.121	14.333	.000



<Figure 5> Interaction Effects of Face on Flow



<Figure 6> Interaction Effects of Face on Enjoyment

teraction effect between the presence of the face and the voice type (F = 14.333, p < .001) was also found to be significant. <Figure 6> is a graph showing the interaction effect between the presence or absence of a face and the voice type, and it can be seen that the user's interactional enjoyment in the AI agent increased when the AI agent has a face and a paralanguage voice.

The results of analyzing the interaction effect between the presence of the AI agent's face and the voice type on the user's satisfaction are shown in <Table 11>. As shown in <Table 11>, the difference in user satisfaction according to the presence or absence of a face (F = 121.793, pp < .001) and the type of voice (F = 133.468, p < .001) is examined. It can be seen that all have a significant effect on the user's satisfaction, and the interaction effect between the presence of the face and the voice type (F = 16.227, p < .001) was also found to be significant. <Figure 7> is a graph showing the interaction effect between the presence or absence of a face and the voice type, and it can be seen that the user's satisfaction with the AI agent increases when the AI agent has a face and a paralanguage voice.

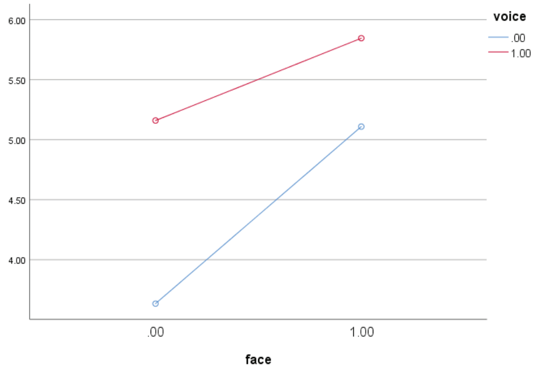
The results of analyzing the interaction effect between the presence of the AI agent's face and the voice type on the user's intention to use are shown in <Table 12>. As shown in <Table 12>, the difference in the user's intention to use according to the presence or absence of a face (F = 110.602, p < .001) and the type of voice (F = 143.827, p < .001) is examined. It can be seen that all have a significant effect on the user's intention to use, and the interaction effect between the presence of the face and the voice type (F = 19.375, p < .001) was also found to be significant. <Figure 8> shows the interaction effect between the presence or absence of a face and the voice type, and it can be seen that the user's intention to use the AI agent increased when the AI agent has a face and a paralanguage voice.

<Table 11> Two-way ANOVA (User Satisfaction)

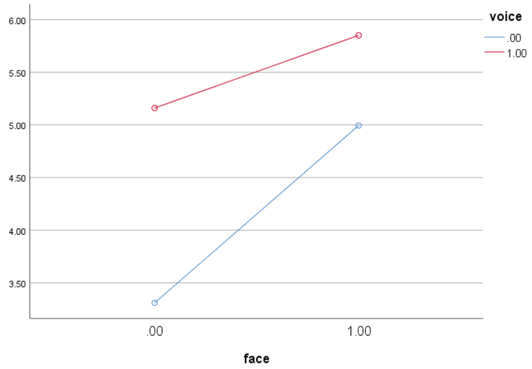
	SS	df	MS	F	p
Face	196.085	1	196.085	121.793	.000
Voice	214.881	1	214.881	133.468	.000
Face*voice	26.125	1	26.125	16.227	.000

<Table 12> Two-way ANOVA (Intention to Use)

	SS	df	MS	F	p
Face	236.906	1	236.906	110.602	.000
Voice	308.073	1	308.073	143.827	.000
Face*voice	41.501	1	41.501	19.375	.000



<Figure 7> Interaction Effects of Face on Satisfaction



<Figure 8> Interaction Effects of Face on Intention

4.2. Study2: Structural Modeling

4.2.1. Verification of the Validity and Reliability of Measurement Tools

In this study, the measurement items for the variables were derived from prior research, and an exploratory factor analysis was performed to validate the construct validity, as shown in <Table 13>. The varimax method was employed for factor rotation, and the factor analysis results revealed no items with low factor loadings below .50 or high factor loadings above .50 for multiple factors. Six factors were extracted as anticipated: factor 1 was identified as 'flow,' factor 2 as 'trust,' factor 3 as 'intimacy,' factor 4 as 'interactional enjoyment,' factor 5 as 'user satisfaction,' and factor 6 as 'intention to use.' An analysis of the factor loadings for each variable's measurement items demonstrated that all items exhibited high values equal to or greater than .50, thereby confirming the construct validity.

The evaluation of construct reliability (CR) and average variance extracted (AVE) was conducted to assess the convergent validity of the research variables. Convergent validity represents the correlation level among two or more measurement items for a single latent variable. Generally, if a variable's CR is at least .70 and the AVE is .50 or higher, the variable is considered to possess convergent validity. As shown in <Table 14>, all research variables—including flow (.941), trust (.963), intimacy (.935), interactional enjoyment (.959), user satisfaction (.952), and intention to use (.941)—displayed exceptionally high values exceeding .90 for CR. Similarly, for AVE, all research variables—flow (.800), trust (.788), intimacy (.828), interactional enjoyment (.855), user satisfaction (.833), and intention to use (.889)—exhibited values above .50, confirming

<Table 13> Factor Loadings of the Measurement Items of Individual Variables (Exploratory Factor Analysis)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Flow1	0.903					
Flow2	0.881					
Flow3	0.887					
Flow4	0.906					
Trust1		0.776				
Trust2		0.908				
Trust3		0.884				
Trust4		0.917				
Trust5		0.926				
Trust6		0.902				
Trust7		0.892				
Intimacy1			0.921			
Intimacy2			0.879			
Intimacy3			0.930			
Interactional enjoyment 1				0.931		
Interactional enjoyment2				0.922		
Interactional enjoyment3				0.92		
Interactional enjoyment4				0.926		
User satisfaction1					0.915	
User satisfaction2					0.931	
User satisfaction3					0.899	
User satisfaction4					0.906	
Intention to use1						0.953
Intention to use2						0.933

<Table 14> Verification of the Validity and Reliability of Measurement Tools for Study Variables

Constructs	AVE	CR	Cronbach's alpha
Flow	.800	.941	.916
Trust	.788	.963	.955
Intimacy	.828	.935	.910
Interactional enjoyment	.855	.959	.943
User satisfaction	.833	.952	.933
Intention to use	.889	.941	.943

convergent validity.

Subsequently, Cronbach's α values were calculated to verify the reliability of the measurement items composing the research variables. Cronbach's α values are generally deemed reliable when they fall within the range of 0.6 to 0.7. The analysis results showed that all research variables, including flow (.916), trust (.955), intimacy (.910), interactional enjoyment (.943), user satisfaction (.933), and intention to use (.943), had exceptionally high Cronbach's α values, not lower than .90. This indicates that the reliability of the measurement instruments for evaluating all research variables was ensured.

Lastly, an assessment of the discriminant validity among latent variables will be conducted. Discriminant validity indicates the degree to which one latent variable is distinct from another. As per the evaluation method, discriminant validity exists between two latent variables if the square root of each variable's AVE is higher than their correlation coefficient (Barclay et al., 1995). Discriminant validity was assessed by comparing the correlation coefficient presented in <Table 15> to the square root of the AVE. The outcome revealed that the correlation coefficient (.919) for the variables flow and intimacy, which had the highest correlation, was higher than the square root of the AVE. However, the correlation coefficients between all other variables were lower than the square root of the AVE, ensuring discriminant validity for the overall analysis.

The PLS program was utilized to evaluate the research hypotheses and investigate the structural relationships between the study variables. The outcomes of the tests for the research hypotheses, which were designed to examine the structural connections among variables such as flow, trust, intimacy, interactional enjoyment, user satisfaction, and intention to use, are presented in <Table 16> and <Figure 9>.

Initially, research hypotheses 1, 2, 3, and 4, which postulated that the display of the agent's face would positively impact the user's flow, trust, intimacy, and interactional enjoyment, were examined. The findings demonstrated that the presence of the agent's face significantly and positively influenced the user's flow (path coefficient = .336, t = 11.370), trust (path coefficient = .316, t = 10.210), intimacy (path coefficient = .285, t = 9.130), and interactional enjoyment (path coefficient = .335, t = 10.536). Consequently, research hypotheses 1, 2, 3, and 4 were accepted.

Next, study hypotheses 5, 6, 7, and 8, which predicted that the paralanguage as the agent's voice would have positive (+) effects on the user's flow, trust, intimacy, and interactional enjoyment, were

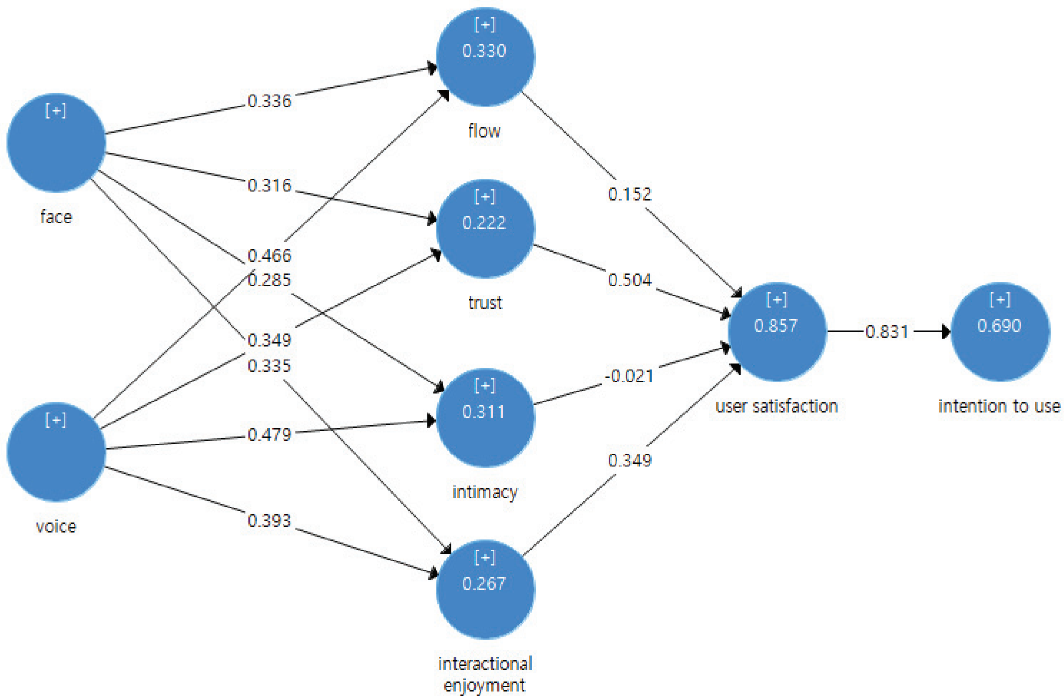
<Table 15> Verification of the Discriminant Validity of Measurement Tools for Study Variables

	Flow	Trust	Intimacy	Interactional enjoyment	User satisfaction	Intention to use
Flow	0.894					
Trust	0.828	.887				
Intimacy	0.919	0.815	.909			
Interactional enjoyment	0.871	0.75	0.794	.924		
User satisfaction	0.793	0.783	0.771	0.682	.912	
Intention to use	0.854	0.831	0.86	0.732	0.879	.942

<Table 16> Results of Hypothesis Tests

H	Path	Path coefficient	t-value	Result
H1	Face → Flow	.336	11.370***	Sig.
H2	Face → trust	.316	10.210***	Sig.
H3	Face → intimacy	.285	9.130***	Sig.
H4	Face → interactional enjoyment	.335	10.536***	Sig.
H5	voice → Flow	.466	16.091***	Sig.
H6	voice → trust	.349	10.682***	Sig.
H7	voice → intimacy	.479	16.771***	Sig.
H8	voice → interactional enjoyment	.393	12.218***	Sig.
H9	Flow → user satisfaction	.152	2.308*	Sig.
H10	Trust → user satisfaction	.504	12.406***	Sig.
H11	Intimacy → user satisfaction	-.021	.449	-
H12	interactional enjoyment → user satisfaction	.349	5.642**	Sig.
H13	user satisfaction → intention to use	.831	46.105***	Sig.

Note: * p < 0.05 ** p < 0.01 *** p < 0.001



<Figure 9> Results of Verification of the Study Model

tested. The results indicated that the paralanguage had significant positive (+) effects on all of the user's flow (path coefficient = .466, t = 16.091), trust (path coefficient = .349, t = 10.682), intimacy (path coefficient = .479, t = 16.771), and interactional enjoyment (path coefficient = .393, t = 12.218). Therefore, study hypotheses 5, 6, 7, and 8 were adopted.

Next, study hypotheses 9, 10, 11, and 12, which predicted that user's flow, trust, intimacy, and interactional enjoyment would have positive (+) effects on user satisfaction, were tested, and the results indicated that user's flow (path coefficient = .152), t = 2.308), trust, (path coefficient = .504, t = 12.406), and interactional enjoyment (path coefficient = .349, t = 5.642) had significant positive (+) effects on user satisfaction. However, the results indicated that the user's intimacy (path coefficient = -.021, t = .449) did not significantly affect user satisfaction. Therefore, hypotheses 9, 10, and 12 were adopted, but study hypothesis 11 was rejected.

Finally, study hypothesis 13, which predicted that user satisfaction would have a positive (+) effect on the intention to use, was tested, and the results indicated that user satisfaction (path coefficient = .349, t = 5.642) had a significantly positive (+) effect on the intention to use. Therefore, research hypothesis 13 was adopted.

Ⅴ. Discussions and Conclusion

This study embarked on an empirical exploration into the intricate relationships between user experience factors such as flow, trust, intimacy, and interactional enjoyment. By dissecting these dynamics across four distinct experimental groups (A, B, C, and D), I delved into how the presence or absence of an agent's face and the type of voice used—paralanguage versus computer-generated—impact these critical user experience metrics.

5.1. Study 1: Effects of Agent Face Exposure and Voice Type on User Satisfaction

The findings from Study 1 highlight the considerable positive influence that the exposure of the agent's face has on all measured aspects of the user experience. This underscores the human tendency to respond more favourably to interfaces that mimic human interaction, as evidenced by increased engagement and satisfaction when interacting with a face, compared to text-based interfaces. The implications of these results are significant, suggesting that the integration of human-like faces in AI services could markedly enhance user satisfaction and engagement.

The elucidation of the findings engenders profound insights into the dynamics of human-AI interaction, highlighting the quintessential role of agent face exposure and voice type in augmenting user engagement and satisfaction. By dissecting these dynamics across four distinct experimental groups (A, B, C, and D), I delved into how the presence or absence of an agent's face and the type of voice used—paralanguage versus computer-generated—impact these critical user experience metrics.

5.1.1. Agent Face Exposure: A Facet of Human-AI Interaction Enhancement

The empirical evidence underscores the pivotal role of agent face exposure in significantly augmenting user flow, engendering trust, fostering intimacy, and amplifying interactional enjoyment. This phenomenon elucidates that the incorporation of visual presence within AI voice services transcends the conventional text-based interfaces, crafting a more immersive and fulfilling user experience. The integration of a human-like face within the AI interface ostensibly nurtures a more robust connection and trust between the user and the AI entity. This is consonant with antecedent research (Sproull et al., 1996; Walker et al., 1994), which postulates that interfaces embellished with facial representations are perceived as more congenial and gratifying. Consequently, it is discernible that the exposure of the agent's human-like face significantly elevates the user's level of flow, trust, intimacy, and interactional enjoyment vis-à-vis scenarios where such exposure is absent.

5.1.2. Voice Type: Enhancing User Experience with Paralanguage Voice

Moreover, the type of voice employed by the AI agent, particularly its paralinguistic features, emerges as a critical determinant impacting the aforementioned dimensions of user experience. This insight delineates that a voice imbued with human-like tonalities and emotional undertones substantially enriches the user's flow, trust, intimacy, and interactional enjoyment within the context of AI service engagement. Ergo, it is observable that the deployment of a paralinguistic voice by the agent, as opposed to a mechanistic sonority, elevates the user's experience across these metrics.

In summation, the findings illuminate the indispensable role of agent face exposure and paralinguistic voice characteristics in enhancing the depth and quality of user engagement in AI interfaces. These insights contribute to the broader discourse on human-computer interaction, underscoring the importance of designing AI agents that emulate human-like attributes to foster a more natural and satisfying user experience.

5.2. Study 2: Hypothesis Tests and Results

In Study 2, The primary findings, focused on the study's hypotheses, are summarized below:

First, the effects of exposure and non-exposure of the agent's face on the user's flow, trust, intimacy, and interactional enjoyment were analyzed, and according to the results, face exposure was identified to have significant positive (+) effects on the user's flow, trust, intimacy, and interactional enjoyment. These results mean that in the voice service combined with artificial intelligence, the case where the human face is exposed can increase the user's flow, trust, intimacy, and interactional enjoyment more than the case where the human face is not exposed. Users were more engaged when interacting with a talking face than with text-based interfaces and spent more time with the face agent (Walker et al., 1994). Interfaces with a face were found to be more satisfying and natural to use (Sproull et al., 1996).

Second, the effects of the agent's voice type (paralanguage and computer-generated voice) on the user's flow, trust, intimacy, and interactional enjoyment were analyzed. Paralanguage was found to have significant positive (+) effects on these factors, indicating that using a paralanguage voice in AI-integrated voice services can enhance the user's flow, trust, intimacy, and interactional enjoyment.

Finally, the effect of user satisfaction on the intention to use was analyzed, and the results indicated that user satisfaction had significant positive (+) effects on the intention to use. Therefore, to increase the user's intention to use, it is necessary to increase user satisfaction. The study also highlighted the strong positive relationship between user satisfaction and the intention to use AI voice services. This underscores the importance of enhancing user satisfaction to increase intention to use, suggesting that both the visual and auditory elements of AI agents play crucial roles in user retention and engagement.

This study examined the effects of face exposure and voice type on user satisfaction with AI-integrated voice services. AI is a critical technology in the fourth industrial revolution, and many countries and companies actively support its development. As a result, companies are focusing on IoT as the next paradigm-shifting product. Apple, Microsoft, and Samsung are working to increase customer loyalty and market share by incorporating AI assistant services in their products, aiming to dominate the voice interface standard in the future.

I further confirmed the positive effects of using paralanguage in voice services on user experience. This aspect of the study highlights the importance of integrating more human-like, nuanced voice responses in AI applications to enhance users' sense of connection and satisfaction. This is demonstrated through experiments that show agents with human-like faces attract more attention, increase engagement, and provide joy and trust in interactions compared to voice-only services like Siri or Bixby. Additionally, it is proven that not only can emotions be conveyed through non-mechanical voices, expressing the full range of human emotions, but speech attributes such as speed, pitch, stress, duration, and pauses can also significantly enhance the delivery and persuasiveness of communication, aligning with the human communication context. This reinforces the CASA (Computers Are Social Actors) paradigm. According to Mehrabian's rule, voice accounts for 38% of communication, and non-verbal messages account for 55%, indicating that in the context of interactions with conversational agents (CAs), theories of human communication can also be applied, accounting for 93% of communication. This underscores the practical significance and implications of this study, highlighting its relevance in applying human communication theories to interactions with conversational agents.

5.3. Implications and Conclusion

Implications of the findings from 1, 2 studies and how they contribute to the understanding of the research topic. The findings have profound practical implications for developers and designers in the AI industry. As AI continues to evolve and become more embedded in our daily lives, understanding and implementing elements that significantly improve user experience are paramount. The study demonstrated that the presence of an agent's face and the use of paralanguage voices can enhance user flow, trust, intimacy, and interactional enjoyment and offers a clear directive for AI developers to humanize AI agents. The study contributes to the understanding of how visual and auditory elements of AI agents influence user experience. These insights have profound practical implications for developers and designers in the AI industry, emphasizing the need for AI design to consider not only the functional aspect, but to give more emphasis on the emotional experience as well.

Developers should consider these findings as a mandate to integrate human-like faces and nuanced, expressive voices in AI interfaces. This approach not only meets the functional requirements of AI applications but also addresses the emotional and social needs of users, fostering a more profound sense of connection and satisfaction.

Moreover, the strong link between user satisfaction and the intention to use suggests a strategic focus on enhancing satisfaction could be a critical driver for user retention and loyalty. This calls for a holistic design strategy that encompasses not just the technical functionalities but also the emotional and social dimensions of user interactions with AI agents.

The recognition of AI agents as social actors shifts the paradigm from viewing AI as mere tools to recognizing them as social actors. This perspective encourages the development of AI agents that are not only efficient and effective but also capable of meaningful social interaction and emotional connection.

In conclusion, this study provides actionable insights for the AI development community, emphasizing the need for a user-centric approach in AI design that considers both the functional and emotional dimensions of user experience. By integrating human-like features such as faces and paralanguage voices, developers can significantly enhance the overall user experience, fostering greater satisfaction, engagement, and loyalty towards AI services. Continuous iteration, informed by user feedback and cross-cultural considerations, will be essential as we advance in creating more personalized and emotionally resonant AI agents. By focusing on personalization and personification, developers can create more engaging, satisfying, and effective AI voice services that meet user needs and expectations, driving the future of AI in the fourth industrial revolution.

<References>

[1] Adam, M., Wessel, M., and Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic Markets, 31*(2), 427-445. https://doi.org/10.1007/s12525-020-00414-7

[2] Agarwal, R., and Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly, 24*(4), 665-694.

[3] Al-Azzawi, A. (2014). *Experience with Technology: Dynamics of User Experience with Mobile Media Devices (Springer Briefs in Computer Science).* London: Springer.

[4] Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self-scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596-612. https://doi.org/10.1037/0022-3514.63.4.596

[5] Back, S. J., and Lee, Y. J. (2012). Game interface based on voice recognition for smartphone. In *Korea Information Technology Association, Proceedings of KIIT Summer Conference.*

[6] Bartel, C. A., and Saavedra, R. (2000). The collective construction of work group moods. *Administrative Science Quarterly, 45*, 197-231. https://doi.org/10.2307/2667070

[7] Bickmore, T. W. (2003). *Relational Agents: Effecting Change Through Human-computer Relationships* (Doctoral Dissertation). Massachusetts Institute of Technology, Boston, MA.

[8] Bickmore, T. W., Fernando, R., Ring, L., and Schulman, D. (2010). Empathic touch by relational agents. *IEEE Transactions on Affective Computing, 1*(1), 60-71. https://doi.org/10.1109/T-AFFC.2010.4

[9] Bickmore, T., and Cassell, J. (2001). Relational agents: A model and implementation of building user trust. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 01* (pp. 396-403).

[10] Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies, 59*(1), 119-155. https://doi.org/10.1016/S1071-5819(03)00018-1

[11] Brill, T. M., Munoz, L., and Miller, R. J. (2022). Siri, Alexa, and other digital assistants: A study of customer satisfaction with artificial intelligence applications. In *The Role of Smart Technologies in Decision Making* (pp.1075-1084). Routledge.

[12] Cacioppo, J. T., and Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219-232.

[13] Carroll, B. A., and Ahuvia, A. C. (2006). Some antecedents and outcomes of brand love. *Marketing Letters, 17*(2), 79-89. https://doi.org/10.1007/s11002-006-4219-2

[14] Chandler, J., and Schwarz, N. (2010). Use does not wear ragged the fabric of friendship: Thinking of objects as a live makes people less willing to replace them. *Journal of Consumer Psychology, 20*(2), 138-145. https://doi.org/10.1016/j.jcps.2009.12.008

[15] Chin, J., Diehl, V., and Norman, K. N. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *International Conference on Human Factors in Computing Systems*.

[16] Coeckelbergh, M. (2011). Humans, animals, and robots. *International Journal of Social Robotics, 3*(2), 197-204.

[17] Coursaris, C. K., and Liu, M. (2009). An analysis of social support exchanges in online HIV/AIDS help groups. *Computers in Human Behavior, 25*(4), 911-918. https://doi.org/10.1016/j.chb.2009.03.006

[18] Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of Humanistic Psychology, 15*(3), 41-63. https://doi.org/10.1177/002216787501500306

[19] Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. *Journal of Leisure Research, 24*(1), 93-94. https://doi.org/10.1080/00222216.1992.11969876

[20] Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience: Steps Toward Enhancing the Quality of Life*. New York: Harper Collins Publishers.

[21] Csikszentmihalyi, M. (1999). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.

[22] Coursaris, C. K., and Liu, M. (2009). An analysis of social support exchanges in online HIV/AIDS self-help groups. *Computers in Human Behavior, 25*(4), 911-918. https://doi.org/10.1016/j.chb.2009.03.006

[23] Daft, R. L., and Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554-571. https://doi.org/10.1287/mnsc.32.5.554

[24] Dash, B., and Davis, K. (2022). Significance of nonverbal communication and paralinguistic featurs in communication: A critical analysis. *International Journal for Innovative Research in Multidisciplinary Field, 8*(4), 172-179.

[25] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340. https://doi.org/10.2307/249008

[26] Davis, R., and Wong, D. (2007). Conceptualizing and measuring the optimal experience of the e-learning environment. *Decision Sciences Journal of Innovative Education, 5*(1), 97-126. https://doi.org/10.1111/j.1540-4609.2007.00129.x

[27] Dinev, T., and Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*(1), 61-80. https://doi.org/10.1287/isre.1060.0080

[28] Dolganov, A. G., and Letnev, K. Y. (2020). Informative modeling of subjective reality for intellectual anthropomorphic robots. *Proceedings of the Materials Science and Engineering, 966*(1), 012084. https://doi.org/10.1088/1757-899X/966/1/012084

[29] Ene, I., Pop, M. I., and Nistoreanu, B. (2019). Qualitative and quantitative analysis of consumers perception regarding anthropomorphic AI designs. In *Proceedings of the International Conference on Business Excellence, 13*(1), 707-716. https://doi.org/10.2478/picbe-2019-0063

[30] Gambino, A., Fox, J., and Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*, 71-85. https://doi.org/10.30658/hmc.1.5

[31] Gao, Y., Pan, Z., Wang, H., and Chen, G. (2018). Alexa, my love: Analyzing reviews of Amazon Echo.

In *2018 IEEE Smart World, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (Smart World/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)* (pp. 372-380). https://doi.org/10.1109/SmartWorld.2018.00094

[32] Ginossar, T. (2008). Online participation: A content analysis of differences in utilization of two online cancer communities by men and women, patients and family members. *Health, 23*(1), 1-12. https://doi.org/10.1080/10410230701697100

[33] Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T., and Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition, 184*, 28-43. https://doi.org/10.1016/j.cognition.2018.12.005

[34] Ho, L. A., and Kuo, T. H. (2010). How can one amplify the effect of e-learning? An examination of high-tech employees' computer attitude and flow experience. *Computers in Human Behavior, 26*(1), 23-31. https://doi.org/10.1016/j.chb.2009.07.007

[35] Hoffman, D. L., and Novak, P. T. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *The Journal of Marketing, 60*(3), 50-68. https://doi.org/10.2307/1251841

[36] Hussein, Z. (2015). Explicating students' behaviours of e-learning: A viewpoint of the extended technology acceptance. *International Journal of Management and Applied Science, 1*(10), 68-73.

[37] Hou, J., and Lee, K. M. (2011, October). Effects of self-conscious emotions on affective and responses in HCI and CMC. [Paper presentation] SIGDOC&11, Italy.

[38] Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly, 37*(1), 81-88. https://doi.org/10.1080/0267257X.2019.1687571

[39] Jiang, Z., and Benbasat, I. (2007). Investigating the influence of the functional mechanisms of online product presentations. *Information Systems Research, 18*(4), 454-470.

[40] Johnson, W. L., Rickel, J. W., and Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*(1), 47-78.

[41] Kahai, S. S., and Cooper, R. B. (2003). Exploring the core concept of media richness theory: The impact of cue multiplicity and feedback immediacy on decision quality. *Journal of Management Information Systems, 20*(1), 263-299. https://doi.org/10.1080/07421222.2003.11045754

[42] Kelshaw, T. (2016, October 5). AI & gender: A Maxus survey, Retrieved from http://www.maxusglobal.com/blog/ai-gender-maxus-survey

[43] Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research, 13*(2), 205-223. http://www.jstor.org/stable/23011056

[44] Lee, D. K., and Borah, P. (2020). Self-presentation on Instagram and friendship development among young adults: A moderated mediation model of media richness, perceived functionality, and openness. *Computers in. Human. Behaviour, 103*, 57-66. https://doi.org/10.1016/j.chb.2019.09.017

[45] Lee, S. Y., and Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies, 103*, 95-105. https://doi.org/10.1016/j.ijhcs.2017.02.005

[46] Lee, U. K. (2018). Accessing the acceptance of the speech recognition based virtual assistant service: Applying UTAUT model. *Journal of Product Research, 38*(5), 111-120. https://doi.org/10.1371/journal.pdig.0000510

[47] Lee, J. R., and Nass, C. I. (2010). Trust in Computers: The Computers-Are-Social-Actors (CASA) Paradigm and Trustworthiness Perception in Human-Computer Communication. *In D. Latusek and A. Gerbasi (Eds.), Trust and Technology in a Ubiquitous*

Modern Environment: Theoretical and Methodological Perspectives (pp. 1-15). IGI Global. https://doi.org/10.4018/978-1-61520-901-9.ch001

[48] Leigh, T., and Summers, J. O. (2013). An initial evaluation of industrial buyers' impressions of salespersons' nonverbal cues. *Journal of Personal Selling and Sales Management, 22*(1), 41-53. https://doi.org/10.1080/08853134.2002.10754292

[49] Lim, W. M., Kumar, S., Verma, S., and Chaturvedi, R. (2022). Alexa, what do we know about conversational commerce? Insights from a systematic literature review. *Psychology & Marketing, 39*(6), 1129-1155.

[50] Littlejohn, S. W., and Foss, K. A. (2010). *Theories of Human Communication* (10th ed.). Waveland Press.

[51] Lukyanenko, R., Maass, W., and Storey, V.C. (2022). Trust in artificial intelligence: From a foundational trust framework to emerging research opportunities. *Electron Markets, 32*, 1993-2020. https://doi.org/10.1007/s12525-022-00605-4

[52] Maxham, J. G. III. (2001). Service recovery's influence on consumer satisfaction, positive word-of-mouth, and purchase intentions. *Journal of Business Research, 54*(1), 11-24. https://doi.org/10.1016/S0148-2963(00)00114-4

[53] Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20*(3), 709-734. https://doi.org/10.2307/258792

[54] Mehrabian, A. (1969). Significance of posture and position in the communication of attitude and status relationships. *Psychological Bulletin, 71*(5), 359-372. https://doi.org/10.1037/h0027349

[55] Mehrabian, A. (1971). Silent messages (Vol. 8, No. 152, p. 30). Belmont, CA: Wadsworth.

[56] Minsky, M. (2007), *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster, New York.

[57] Mhatre, N., Motani, K., Shah, M., and Mali, S. (2016). Donna interactive chat-bot acting as a personal assistant. *International Journal of Computer Applications, 140*(10), p.6-11. https://doi.org/10.5120/ijca2016909460.

[58] Moon, Y., and Nass, C. (1996). How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research, 23*(6), 651-674. https://doi.org/10.1177/009365096023006002

[59] Moorman, C., Deshpande, R., and Zaltman, G. (1993). Factors affecting trust in market research relationships. The Journal of Marketing, 57, 81-101. https://doi.org/10.2307/1252059

[60] Morgan, R., and Hunt, S. (1994). The Commitment-Trust Theory of Relationship Marketing. *The Journal of Marketing, 58*, 20-38. http://dx.doi.org/10.2307/1252308

[61] Mori, M. (1970). The uncanny valley. *Energy, 7*(4), 33-35

[62] Morkes, J., Kernal, H. K., and Nass, C., (1998). Humor in task-orientated computer-mediated communication and human computer interaction. *Conference on Human Factors in Computing Systems.* Los Angeles.

[63] Nass, C., and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81-103.

[64] Novak, T. P., Hoffman, D. L., and Yung, Y. F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science, 19*(1), 22-42. http://www.jstor.org/stable/193257

[65] Nowak, K. L., and Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer- Mediated Communication, 11*(1), 153-178. https://doi.org/10.1111/j.1083-6101.2006.tb00308.x

[66] Nowak, K. L., and Rauh, C. (2008). Choose your "buddy icon" carefully: The influence of avatar androgyny, anthropomorphism and credibility in online interactions. *Computers in Human Behavior, 24*(4), 1473-1493. https://doi.org/10.1016/j.chb.2007.

05.005
[67] Orr, D. A., and Sanchez, L. (2018). Alexa, did you get that? Determining the evidentiary value of data stored by the Amazon Echo. *Digital Investigation, 24*, 72-78. https://doi.org/10.1016/j.diin.2017.12.002.
[68] Park, J., Son, H., Lee, J., and Choi, J. (2018). Driving assistant companion with voice interface using long short-term memory networks. *Transactions on Industrial Informatics, 15*(1), 582-590.
[69] Pelau, C., Volkmann, C., Barbul, M., and Bojescu, I. (2023). The role of attachment in improving consumer-AI interactions. *Proceedings of the International Conference on Business Excellence, 17*, 1075-1084. https://doi.org/10.2478/picbe-2023-0097
[70] Peters, D., Calvo, R. A., and Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00797
[71] Picard, R. W. (2003). Affective computing: challenges. *International Journal of Human Computer Studies, 59*(1-2), 55-64. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S1071581903000521
[72] Picard, R. W. (2008). Toward machines with emotional intelligence. In G. Matthews, M. Zeidner, and R. D. Roberts (Eds.), *The Science of Emotional Intelligence: Knowns and Unknowns, Series in Affective Science*. New York: Oxford Academic. https://doi.org/10.1093/acprof:oso/9780195181890.003.0016.
[73] Qiu, L., and Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems, 25*(4), 145-182. https://doi.org/10.2753/MIS0742-1222250405
[74] Rafaeli, S. (1988). Interactivity: From new media to communication. In R. P. Hawkins, J. M. Wiemann, and S. Pingree (Eds.), *Sage Annual Review of Communication Research: Advancing Communication Science* (Vol. 16, pp. 110-134). Beverly Hills, CA: Sage. https://doi.org/10.1111/j.1083-6101.1997.tb00201.x
[75] Reeves, B., and Nass, C. I. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Center for the Study of Language and Information; Cambridge University Press.
[76] Rempel, J., Holmes, J., and Zanna, M. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49*, 95-112. https://psycnet.apa.org/doi/10.1037/0022-3514.49.1.95
[77] Ruel, H., and Njoku, E. (2021), AI redefining the hospitality industry. *Journal of Tourism Futures, 7*(1), 53-66. https://doi.org/10.1108/JTF-03-2020-0032
[78] Song, Y., and Luximon, Y. (2020). Trust in AI agent: A systematic review of facial anthropomorphic trustworthiness for social robot design. *Sensors, 20*(18), 5087.
[79] Suh, K. S., and Lee, Y. E. (2005). The effects of virtual reality on consumer learning: An empirical investigation. *MIS Quarterly, 29*(4), 673-697.
[80] Sundar, S. S., and Kim, J. (2005). Interactivity and persuasion: Influencing attitudes with information and involvement. *Journal of Interactive Advertising, 5*(2), 5-18.
[81] Sproull, L., Subramani, M., Kiesler, S., Walker, J. H., and Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11*(2), 97-124. https://doi.org/10.1207/s15327051hci1102_1
[82] Littlejohn, S. W., and Foss, K. A. (Eds.) (2009). *Encyclopedia of Communication Theory*. Thousand Oaks, Calif: Sage.
[83] Tepper, D. T., and Haase, R. F. (1978). Verbal and nonverbal communication of facilitative conditions. *Journal of Counseling Psychology, 25*(1), 35-44. https://doi.org/10.1037/0022-0167.25.1.35
[84] Van der Heijden, H. (2003). Factors influencing the usage of websites: The case of a generic portal in The Netherlands. *Information & Management, 40*, 541-549. https://doi.org/10.1016/S0378-7206(02)00079-4.

[85] Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly, 28*(4), 695-704. https://doi.org/10.2307/25148660

[86] Wainwright, M. J. (1999). Visual adaptation as optimal information transmission. *Vision Research, 19*(23), 3960-3974. https://doi.org/10.1016/S0042-6989(99)00101-7.

[87] Wang, W., and Benbasat, I. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems, 6*(3), 72-101.

[88] Walker, J., Sproull, L., and Subramani, M. (1994). Using a human face in an interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 85-91). https://doi.org/10.1145/259963.260290.

[89] Waytz, A., Cacioppo, A., and Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219-232.

[90] Zanbaka, C., Goolkasian, P., and Hodges, L. (2006). Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In *CHI 2006*(pp. 1153-1162). https://doi.org/10.1145/1124772.1124945

[91] Zhu, L., Benbasat, I., and Jiang, Z. (2010). Let's shop online together: An empirical investigation of collaborative online shopping support. *Information Systems Research, 21*(4), 872-891. https://doi.org/10.1287/isre.1080.0218

◆ About the Authors ◆

SeoYoung Lee

SeoYoung Lee received the Ph.D. degree at Yonsei University, Seoul and a Master of Communication also from Yonsei University, Seoul, Korea. Now she is working as a full-time professor at DongYang University, North Seoul Campus. She was registered in “Marquis Who's Who in the world” and received Marquis Albert Nelson Lifetime Achievement Award from 2017-2020. Her research interests include Media, Speech Communication, Information Systems and HCI etc. Her papers have been published in KCI journals and SSCI journals.

Submitted: July 3, 2023; 1st Revision: February 1, 2024; 2nd Revision: March 16, 2024; Accepted: April 24, 2024