Intelligent Virtual Agents

Subjects: Computer Science, Artificial Intelligence
Contributor: Amal Abdulrahman, Deborah Richards

The use of intelligent virtual agents (IVA) to support humans in social contexts will depend on their social acceptability. Acceptance will be related to the human's perception of the IVAs as well as the IVAs' ability to respond and adapt their conversation appropriately to the human. Adaptation implies computer-generated speech (synthetic speech), such as text-to-speech (TTS).

Keywords: embodied conversational agent ; stress management ; voice

1. Introduction

Intelligent virtual agents (IVAs) are artificially intelligent animated characters, designed to mimic natural human–human interaction for various objectives including entertainment [1], education [2], and healthcare [3]. Compared to conversational agents (like chatbots or voice assistants) that do not have a visual representation or embodiment, building acceptable and engaging IVAs for fruitful interactions is more challenging due to the appearance dimension included in the design of IVAs. For this reason, they are also commonly known as embodied conversational agents. The design of IVAs requires multidisciplinary effort, including psychology [4] and artificial intelligence [5] expertise to achieve capabilities such as emotion modeling [6][7]. Yet, these capabilities are largely still under laboratory investigation, and what factors may influence their acceptance by users is an ongoing research question [3][8].

People tend to perceive the presence, and consequently interact socially, with computers in similar ways as they naturally do with each other [9][10]. With the increasing use of IVAs in social, entertainment, health, and work settings, the acceptance of this technology is growing. Great effort is being devoted to increase this acceptance by giving the technology life by simulating the natural look and behavior of a human being (i.e., anthropomorphism) [11]. However, the theory of the "uncanny valley", which was first introduced by Mori [12], predicts that this acceptance will not last, and at some point, IVAs will manifest an unacceptable behavior/appearance and induce a feeling of eeriness and discomfort.

According to Nowak [13], presence has three measurable dimensions: social presence, co-presence, and presence as transportation. Social presence evaluates the ability of the media itself to bring the sense of presence/existence, co-presence evaluates if the sense of presence is established (mutual awareness and understanding), and presence as transportation evaluates the sense of being immersed in the environment as if the user has moved to the virtual world.

This differentiation between social presence and co-presence is important to understand the impact of anthropomorphism. In the literature, it is assumed that increasing the agent's realism (anthropomorphism) leads to higher social presence, and consequently increases the IVA's likeability and acceptance (e.g., [14][15]), with contradictory findings regarding the impact of anthropomorphism on the user–agent relationship and interaction goal (e.g., [16][17][18][19]). However, considering the uncanny valley problem, Brenton et al. [20] suggested that the eeriness problem or "uncanny valley" occurs as a result of how the user perceives the agent's co-presence. Co-presence is defined as having the sense of perceiving the IVA by the user and having the sense of being actively perceived by the IVA as well [13]. This mutual awareness is a conscious process that is developed when an interaction goal is established [13]. This conscious awareness and established goal help the user to adapt to a low level of features (low anthropomorphism) when, for example, using talking objects (agents) and interacting socially with them [21][22]. On the other hand, social presence is defined as having the sense of being 'there', in the same environment with the virtual agent. This clear distinction makes measurement of social presence more appropriate in virtual reality studies, such as [23].

The study of anthropomorphism has included the agent's appearance, facial expressions, voice, spoken responses, and physical movements. Due to the need for speech to be adaptive and the predominant use of synthetic voice/text-to-speech (TTS) technology rather than recorded real human voice, the role of voice anthropomorphism (human vs. synthetic voice), its effect on co-presence, and its impact on the interaction outcome (intentions to change a behavior).

2. Anthropomorphism and Co-Presence

People naturally enter into others' presence with very little information about the interaction situation, and hence they pay more attention to others' appearance and social cues [24]. The same rule applies when people interact with a human being or virtual character [9][25]. Presence has been considered as a major tool to evaluate social virtual agents by measuring to what extent the agent could engage the human user in the virtual world.

According to the ethopoeia theory, adding human social characteristics to an object increases the human's tendency to unconsciously accept the object as a social entity [9]. However, the threshold model of social influence posits that these characteristics should be as real as those of human beings (anthropomorphism) to boost a virtual agent's presence [26]. However, realism and presence are not always in a direct proportional relationship, and there is a point where eeriness is sensed—the uncanny valley problem [20] in human–agent interaction—which has still not been fully explained. Some researchers regard the problem as a subconscious cognitive reaction when the user's expectation is violated [27], while others regarded it as an emotional arousal response to the agent's appearance and behavior, correlated with the agent's presence [28]. Therefore, measuring the agent's presence, specifically co-presence, can reveal when the eeriness occurs as a result of anthropomorphism [20].

Voice is one of the human characteristics that increase the IVA's realism. It was found to be of great influence in applying social rules to objects where people perceived different voices of the same object as different agents [29]. A simple talking object, a tissue box saying "bless you", was perceived as social, as a humanoid robot, and as a human being [22]. Although the participants in [22] rated the robot and the box as eerier and less human than the human being, they rated them as being as social, attractive, friendly, and intelligent as the human being.

Voice anthropomorphism boosts the users' likeability [30]; however, there is a debate on its influence on the final interaction outcome. On one hand, some studies found anthropomorphism contributed towards increasing human–agent understanding, trust, connectedness, and task outcome [16][17][31][32]. On the other hand, other studies reported no influence on the outcome [33][34], or they had concerns with the use of higher anthropomorphism. For example, adding a human tone to the chatbot's voice influenced the user's online shopping behavior. A corporate's formal and emotionless tone, compared to an employee's informal and emotional tone, was perceived to be less credible when making important decisions during shopping, such as when shopping for gifts [35], and to have a higher privacy risk, which impacts the behavioral intention—to register on the business website [18].

3. Anthropomorphism and Congruence

Early works in psychology concluded that humans naturally tend to prefer clarity and consistency in perception and interaction. The principle of grouping (Gestalt theory) postulates that people always perceive and organize surroundings/objects according to certain rules, including similarity [36]. When inconsistency occurs, the human brain gets confused and takes a longer time to respond, which is called the Stroop effect [37]. This confusion is not only limited to perceiving objects but extends to almost everything, including perceiving other humans and their personalities. The auditory Stroop effect was examined by Green and Barber [38], where their study participants found it more difficult to identify the gender of a speaker when a male speaker says 'girl' or a female speaker says 'man'.

The same experience is extended to human–computer and human–robot interaction. Isbister and Nass [4] underlined the importance of the consistency in verbal and non-verbal cues of a virtual character, regardless of its personality or that of the users. Mitchell, Szerszen Sr, Lu, Schermerhorn, Scheutz, and MacDorman [33] reported that users found that a human face with a synthetic voice or a humanoid robot with a human voice caused significantly higher eeriness than when the face and voice were matched. Users rated the mismatched/inconsistent scenarios as eerier than the matched ones and the robot with a synthetic voice as the warmest. Similar results were reported earlier by Gong and Nass [39] using a human face or humanoid face matched with human or computerized voices. Moore [40] proved this phenomenon of the mismatch effect mathematically using a Bayesian model of categorical perception and called it 'perceptual tension'.

The impact of matching the agent's look with its voice (congruency) on user–agent trust has been further studied, with contrary findings. For example, Torre et al. [41] introduced an investment game to 120 participants to play with either a generous or a mean robot that switches from a human to a synthetic voice (or vice versa) in the middle of the game. The results indicate the importance of matching the voice with the look to form the first impression, which impacts the user's trust in the agent. However, other studies endorsed the use of a human recorded voice over a synthetic voice, such as Chérif and Lemoine [42], who concluded that a virtual assistant with a human voice on commercial websites increases the assistant's presence, and the users' behavioral intentions as well (N = 640), but it does not impact the users' trust in the

website. They reported a negative influence of TTS on user–agent trust, which then negatively impacts behavioral intentions.

Further studies failed to find a significant difference between the use of human recorded voices and TTS regarding the system's desired outcome. With 138 undergraduate students, Zanbaka, Goolkasian, and Hodges [32] investigated how different persuasive messages would be perceived by students when the messages were delivered by human beings, IVAs, and a talking cat. The three settings had two versions: male and female. Human voices were used in the three settings. The results showed that participants perceived the IVAs and cats to be as persuasive as human speakers. Further, they showed a similar attitude towards the virtual characters and the human—male participants were more persuaded by real and virtual female speakers, while female participants were more persuaded by real and virtual male speakers. This stereotypical response is in line with the findings from voice-only studies by Mullennix, Stern, Wilson, and Dyson [30]. Similarly, other studies reported no significant differences between voice settings in terms of social reactions [43] or keeping a distance between a user and a robot [44].

The impact of the agent's voice type (human recorded voice vs. synthetic/computerized voice) could depend on context. In a learning scenario, Dickerson et al. [45] reported that participants found synthetic voice unnatural, but they quickly adapted and rated it as being as intelligent as the agent with a human voice. The system outcome was met equivalently in both groups, but they concluded that a human voice is preferred only for expressive virtual agents, as it can convey the speaker's emotions and attitudes. In a voice-only system, Noah et al. [46] concluded that synthetic voice would harm the user's trust in high-risk and highly personal contexts and that the use of human or synthetic voice should be investigated before deploying a system by studying the user's expectations and concerns with the type of the voice.

In a matching task, Torre, Latupeirissa, and McGinn [41] asked 60 participants to match one robot out of eight robots' pictures with a voice from a set of voices that varied in terms of gender and naturalness, for four different contexts: home, school, restaurant, and hospital. The results showed that the more mechanical the look of the robot, the greater the tendency for the robot to be matched with a synthetic voice, and that the matching was significantly different across the contexts.

Black and Lenzo [47] suggested the use of domain-limited (professional vs. general) synthetic voice to replace the use of human voice; however, Georgila et al. [48] reported no significant difference between a general-purpose synthetic voice and a good domain-limited synthetic voice in terms of perceived naturalness, conversational aspects, and likeability.

References

1. Yuan, X.; Chee, Y.S. Design and evaluation of Elva: An embodied tour guide in an interactive virtual art gallery. Comput. Animat. Virtual Worlds 2005, 16, 109–119.
2. Aljameel, S.S.; O'Shea, J.D.; Crockett, K.A.; Latham, A.; Kaleem, M. Development of an Arabic conversational intelligent tutoring system for education of children with ASD. In Proceedings of the 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Paris, France, 26–28 June 2017; pp. 24–29.
3. Provoost, S.; Lau, H.M.; Ruwaard, J.; Riper, H. Embodied conversational agents in clinical psychology: A scoping review. J. Med. Internet Res. 2017, 19, e151.
4. Isbister, K.; Nass, C. Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. Int. J. Hum. Comput. Stud. 2000, 53, 251–267.
5. Diederich, S.; Brendel, A.B.; Kolbe, L.M. Towards a Taxonomy of Platforms for Conversational Agent Design. In Proceedings of the International Conference on Wirtschaftsinformatik, Siegen, Germany, 24–27 February 2019.
6. Clore, G.L.; Ortony, A. Psychological construction in the OCC model of emotion. Emot. Rev. 2013, 5, 335–343.
7. Picard, R.W. Affective Computing; MIT Press: Cambridge, MA, USA, 2000.
8. Vaidyam, A.N.; Wisniewski, H.; Halamka, J.D.; Kashavan, M.S.; Torous, J.B. Chatbots and conversational agents in mental health: A review of the psychiatric landscape. Can. J. Psychiatry 2019, 64, 456–464.
9. Reeves, B.; Nass, C.I. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places; Cambridge University Press: Cambridge, England, 1996.
10. Schultze, U.; Brooks, J.A.M. An interactional view of social presence: Making the virtual other "real". Inf. Syst. J. 2019, 29, 707–737.

11. Van Pinxteren, M.M.; Pluymaekers, M.; Lemmink, J.G. Human-like communication in conversational agents: A literature review and research agenda. J. Serv. Manag. 2020, 31, 203–225.
12. Mori, M. the uncanny valley. Energy 1970, 7, 33–35.
13. Nowak, K. Defining and differentiating copresence, social presence and presence as transportation. In Proceedings of the Presence 2001 Conference, Philadelphia, PA, USA, 21 May 2001; pp. 1–23.
14. Oh, C.S.; Bailenson, J.N.; Welch, G.F. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. Front. Robot. AI 2018, 5, 114.
15. Li, M.; Suh, A. Machinelike or Humanlike? A Literature Review of Anthropomorphism in AI-Enabled Technology. In Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS 2021), Maui, HI, USA, 5–8 January 2021; pp. 4053–4062.
16. Kang, H.; Kim, K.J. Feeling Connected to Smart Objects? A Moderated Mediation Model of Locus of Agency, Anthropomorphism, and Sense of Connectedness. Int. J. Hum. Comput. Stud. 2020, 133, 45–55.
17. Kim, S.; Lee, J.; Gweon, G. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019; pp. 1–12.
18. Xie, Y.; Chen, K.; Guo, X. Online anthropomorphism and consumers' privacy concern: Moderating roles of need for interaction and social exclusion. J. Retail. Consum. Serv. 2020, 55, 102119.
19. Schmitt, A.; Zierau, N.; Janson, A.; Leimeister, J.M. Voice as a contemporary frontier of interaction design. In Proceedings of the European Conference on Information Systems (ECIS), Virtual, 22 May 2021.
20. Brenton, H.; Gillies, M.; Ballin, D.; Chatting, D. The uncanny valley: Does it exist. In Proceedings of the Conference of Human Computer Interaction, Workshop on Human Animated Character Interaction, Las Vegas, NV, USA, 22–27 July 2005.
21. Rothstein, N.; Kounios, J.; Ayaz, H.; de Visser, E.J. Assessment of Human-Likeness and Anthropomorphism of Robots: A Literature Review. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, San Diego, CA, USA, 16–20 July 2020; pp. 190–196.
22. Jia, H.; Wu, M.; Jung, E.; Shapiro, A.; Sundar, S.S. When the tissue box says "Bless You": Using speech to build socially interactive objects. In Proceedings of the CHI '13 Extended Abstracts on Human Factors in Computing Systems, Paris, France, 27 April 2013; pp. 1635–1640.
23. Higgins, D.; Zibrek, K.; Cabral, J.; Egan, D.; McDonnell, R. Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. Comput. Graph. 2022, 104, 116–128.
24. Goffman, E. The Presentation of Self in Everyday Life; Harmondsworth: London, UK, 1978.
25. Nowak, K.L.; Biocca, F. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence in Virtual Environments. Presence Teleoperators Virtual Environ. 2003, 12, 481–494.
26. Blascovich, J.; Loomis, J.; Beall, A.C.; Swinth, K.R.; Hoyt, C.L.; Bailenson, J.N. Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. Psychol. Inq. 2002, 13, 103–124.
27. MacDorman, K.F.; Ishiguro, H. The uncanny advantage of using androids in cognitive and social science research. Interact. Stud. 2006, 7, 297–337.
28. Ciechanowski, L.; Przegalinska, A.; Magnuski, M.; Gloor, P. In the Shades of the Uncanny Valley: An Experimental Study of Human–chatbot Interaction. Future Gener. Comput. Syst. 2019, 92, 539–548.
29. Nass, C.; Steuer, J. Voices, Boxes, and Sources of Messages: Computers and Social Actors. Hum. Commun. Res. 1993, 19, 504–527.
30. Mullennix, J.W.; Stern, S.E.; Wilson, S.J.; Dyson, C.-l. Social perception of male and female computer synthesized speech. Comput. Hum. Behav. 2003, 19, 407–424.
31. de Visser, E.J.; Monfort, S.S.; McKendrick, R.; Smith, M.A.; McKnight, P.E.; Krueger, F.; Parasuraman, R. Almost human: Anthropomorphism increases trust resilience in cognitive agents. J. Exp. Psychol. Appl. 2016, 22, 331.
32. Zanbaka, C.; Goolkasian, P.; Hodges, L. Can a virtual cat persuade you?: The role of gender and realism in speaker persuasiveness. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Gaithersburg, MD, USA, 15–17 March 2006; pp. 1153–1162.
33. Mitchell, W.J.; Szerszen Sr, K.A.; Lu, A.S.; Schermerhorn, P.W.; Scheutz, M.; MacDorman, K.F.J.i.-P. A mismatch in the human realism of face and voice produces an uncanny valley. iPerception 2011, 2, 10–12.

34. Cowan, B.R.; Branigan, H.P.; Obregón, M.; Bugis, E.; Beale, R. Voice Anthropomorphism, Interlocutor Modelling and Alignment Effects on Syntactic Choices in Human-Computer Dialogue. Int. J. Hum. Comput. Stud. 2015, 83, 27–42.
35. Barcelos, R.H.; Dantas, D.C.; Sénécal, S. Watch Your Tone: How a Brand's Tone of Voice on Social Media Influences Consumer Responses. J. Interact. Mark. 2018, 41, 60–80.
36. Smith, B. Foundations of Gestalt Theory; Philosophia Verlag: Munich, Germany, 1988.
37. Stroop, J.R. Studies of interference in serial verbal reactions. J. Exp. Psychol. 1935, 18, 643.
38. Green, E.J.; Barber, P.J. An Auditory Stroop Effect with Judgments of Speaker Gender. Percept. Psychophys. 1981, 30, 459–466.
39. Gong, L.; Nass, C. When a Talking-Face Computer Agent Is Half-Human and Half-Humanoid: Human Identity and Consistency Preference. Hum. Commun. Res. 2007, 33, 163–193.
40. Moore, R.K. A Bayesian Explanation of the 'Uncanny Valley' Effect and Related Psychological Phenomena. Sci. Rep. 2012, 2, 1–5.
41. Torre, I.; Latupeirissa, A.B.; McGinn, C. How context shapes the appropriateness of a robot's voice. In Proceedings of the 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Naples, Italy, 31 August–4 September 2020; pp. 215–222.
42. Chérif, E.; Lemoine, J.-F. Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. Rech. Et Appl. En Mark. (Engl. Ed.) 2019, 34, 28–47.
43. Lee, E.-J. The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. Comput. Hum. Behav. 2010, 26, 665–672.
44. Walters, M.L.; Syrdal, D.S.; Koay, K.L.; Dautenhahn, K.; Te Boekhorst, R. Human approach distances to a mechanical-looking robot with different robot voice styles. In Proceedings of the RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany, 1–3 August 2008; pp. 707–712.
45. Dickerson, R.; Johnsen, K.; Raij, A.; Lok, B.; Stevens, A.; Bernard, T.; Lind, D.S. Virtual patients: Assessment of synthesized versus recorded speech. Stud. Health Technol. Inf. 2006, 119, 114–119.
46. Noah, B.; Sethumadhavan, A.; Lovejoy, J.; Mondello, D. Public Perceptions Towards Synthetic Voice Technology. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 2021, 65, 1448–1452.
47. Black, A.W.; Lenzo, K.A. Limited Domain Synthesis; Carnegie-Mellon University Pittsburgh Pa Inst of Software Research Internat: Pittburgh, PA, USA, 2000.
48. Georgila, K.; Black, A.W.; Sagae, K.; Traum, D.R. Practical Evaluation of Human and Synthesized Speech for Virtual Human Dialogue Systems. In Proceedings of the LREC, Istanbul, Turkey, 21 May 2012; pp. 3519–3526.
