PRESS RELEASE

For immediate release: June 1, 2021

Subject line: *Talking Points*: **The Role of Computer Voice in the Future of Speech-Based Human-Computer Interaction**

(Tokyo, June 1) **In the modern day, our interactions with voice-based devices and services continue to increase. In this light, researchers at Tokyo Institute of Technology and RIKEN, Japan, have performed a meta-synthesis to understand how we perceive and interact with the voice (and the body) of various machines. Their findings have generated insights into human preferences, and can be used by engineers and designers to develop future vocal technologies.**

As humans, we primarily communicate vocally and aurally. We convey not just linguistic information, but also the complexities of our emotional states and personalities. Aspects of our voice such as tone, rhythm, and pitch are vital to the way we are perceived. In other words, the way we say things matters.

With advances in technology and the introduction of social robots, conversational agents, and voice assistants into our lives, we are expanding our interactions to include computer agents, interfaces, and environments. Research on these technologies can be found across the fields of human-agent interaction (HAI), human-robot interaction (HRI), human-computer interaction (HCI), and human-machine communication (HMC), depending on the kind of technology under study. Many studies have analyzed the impact of computer voice on user perception and interaction. However, these studies are spread across different types of technologies and user groups and focus on different aspects of voice.

In this regard, a group of researchers from Tokyo Institute of Technology (Tokyo Tech), Japan, RIKEN Center for Advanced Intelligence Project (AIP), Japan, and gDial Inc., Canada, have now compiled findings from several studies in these fields with the intention to provide a framework that can guide future design and research on computer voice. As lead researcher Associate Professor Katie Seaborn from Tokyo Tech (Visiting Researcher and former Postdoctoral Researcher at RIKEN AIP) explains, "Voice assistants, smart speakers, vehicles that can speak to us, and social robots are already here. We need to know how best to design these technologies to work with us, live with us, and match our needs and desires. We also need to know how they have influenced our attitudes and behaviors, especially in subtle and unseen ways."
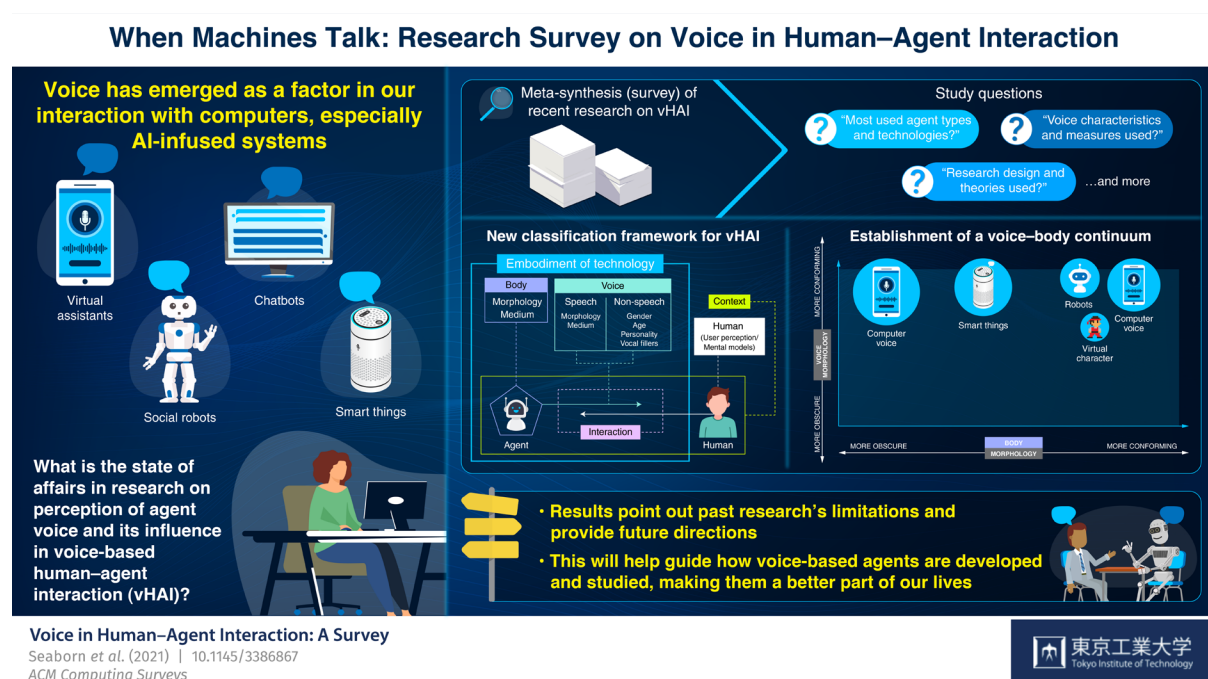
The team's survey considered peer-reviewed journal papers and proceedings-based conference papers where the focus was on the user perception of agent voice. The source materials encompassed a wide variety of agent, interface, and environment types and

technologies, with the majority being "bodyless" computer voices, computer agents, and social robots. Most of the user responses documented were from university students and adults. From these papers, the researchers were able to observe and map patterns and draw conclusions regarding the perceptions of agent voice in a variety of interaction contexts.

The results showed that users anthropomorphized the agents that they interacted with and preferred interactions with agents that matched their personality and speaking style. There was a preference for human voices over synthetic ones. The inclusion of vocal fillers such as the use of pauses and terms like "I mean…" and "um" improved the interaction. In general, the survey found that people preferred human-like, happy, empathetic voices with higher pitches. However, these preferences were not static; for instance, user preference for voice gender changed over time from masculine voices to more feminine ones. Based on these findings, the researchers were able to formulate a high-level framework to classify different types of interactions across various computer-based technologies.

The researchers also considered the effect of the body, or morphology and form factor, of the agent, which could take the form of a virtual or physical character, display or interface, or even an object or environment. They found that users tended to perceive agents better when the agents were embodied and when the voice "matched" the body of the agent.

The field of human-computer interaction, particularly that of voice-based interaction, is a burgeoning one that continues to evolve almost daily. As such, the team's survey provides an essential starting point for the study and creation of new and existing technologies in voice-based human-agent interaction (vHAI). "The research agenda that emerged from this work is expected to guide how voice-based agents, interfaces, systems, spaces, and experiences are developed and studied in the years to come," Prof. Seaborn concludes, summing up the importance of their findings.

When Machines Talk: Research Survey on Voice in Human–Agent Interaction

Voice in Human–Agent Interaction: A Survey
Seaborn *et al.* (2021) | 10.1145/3386867
ACM Computing Surveys

*Corresponding author's email: seaborn.k.aa@m.titech.ac.jp

Contact
Emiko Kawaguchi
Public Relations Division,
Tokyo Institute of Technology
media@jim.titech.ac.jp
+81-3-5734-2975

About Tokyo Institute of Technology
Tokyo Tech stands at the forefront of research and higher education as the leading university for science and technology in Japan. Tokyo Tech researchers excel in fields ranging from materials science to biology, computer science, and physics. Founded in 1881, Tokyo Tech hosts over 10,000 undergraduate and graduate students per year, who develop into scientific leaders and some of the most sought-after engineers in industry. Embodying the Japanese philosophy of "monotsukuri," meaning "technical ingenuity and innovation," the Tokyo Tech community strives to contribute to society through high-impact research.
https://www.titech.ac.jp/english/