Talk the Talk: How Human Conversational Agents Build Trust

John M. Carroll
Center for Human-Computer Interaction
Pennsylvania State University
University Park, PA 16802, USA
jmcarroll@psu.edu

Sooyeon Lee
Center for Human-Computer Interaction
Pennsylvania State University
University Park, PA 16802, USA
sul131@psu.edu

Chun-Hua Tsai
Center for Human-Computer Interaction
Pennsylvania State University
University Park, PA 16802, USA
ctsai@psu.edu

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
Conversational agents (CAs) provide convenient access to everyday information and services by supporting natural language interactions. CAs leverage conversational turn taking, an interaction that is familiar and engaging for humans, and one that allows participants to negotiate and refine messages through a series of turns. We investigated a commercial service interaction in which visually impaired humans are assisted by a remote sighted human agent. This service interaction is broadly applied, and largely conversational. We describe characteristics of human conversational agents as a possible model for achieving more effective computational CAs.

Author Keywords
conversational agents; human conversational agents; conversations for vision; visual prosthetics

CCS Concepts
•**Human-centered computing** → **Collaborative and social computing systems and tools;** *Empirical studies in accessibility; Empirical studies in interaction design;*

Introduction
Conversational agents (CA) provide convenient access to everyday information and services by supporting natural language interactions. A well-known and pioneering exam-

ple is ELIZA, Weizenbaum's rule-based interaction model of a Rogerian psycho-therapeutic interview [3]. This conversational interaction is a clever choice: Rogerian therapists reflect a patient's statements back to them in order to evoke further elaboration. For example, if the patient says, "I feel my boyfriend tries to control me.", the therapist might say "Tell me more about your boyfriend." or "What makes you fell your boyfriend tries to control you." The rules are pretty simple. And people experienced this CA as quite compelling and human-like (note that this was more than 50 years ago).

Typical CA applications today include routine customer services calls, airline travel reservation dialogs, and personal digital assistants, such as Siri and Alexa that set alarms/timers/reminders, update the news and weather, and stream music. Conversational agents are still often quite limited in interactions. For example, airline phone-based CA's use a fixed script to elicit information and populate a form (departure airport, arrival airport, departure date, return date, etc.). This is known to be a limitation in human-computer dialog design: Functionally, it is inflexible and error prone (for example, the human may be thinking about and specify the return date first). Relatively specific and low-level inflexibility with respect to conversational initiative can also remind the human of the computer's limitations and undermine trust more broadly.

In this paper, we describe a commercial service interaction in which people with visual impairment are assisted by a remote sighted human agent. The agent and the client/user interact through voice via the Internet; the client also shares a personal camera feed (e.g., from a smartphone) with agent. This service interaction is broadly applied: People with visual impairment can have an agent read labels, signs, even books to them, can get directions as they follow a route, or can get coaching and feedback as they carry out a cooking project or deliver a lecture.

Human conversational agents

In our fieldwork with the National Federation for the Blind during the past 5 years, we learned that this service is a popular prosthetic approach. We collaborated with the most well-known service-provider (Aira.io) to carry out an interview study of agents [1]. In this short paper, we describe characteristics of human conversational agents in this setting as a possible model for conceptualizing and achieving more effective computational CAs.

A simple way to think about Aira.io agents is that they are *eyes that talk*; the agents describe the visual world to people who cannot fully see it. This is a reasonable concept both because it focuses and simplifies what the agent is expected to do, but also because it acknowledges and emphasizes that the client/user is in charge of the interaction. The prosthetic is conversational, but it is conversation about the vision of the visually impaired person; that person identifies the goals, makes the decisions, and evaluates the outcomes.

In our study, we found that agents are much more than passive *eyes that talk*. Clients are not merely wondering what object is in front of them and asking their prosthetic eyes to tell them. Instead, clients and agents carry out extended conversations. They continually negotiate and adjust the level of detail for descriptions of situations: clients react to the agent's description by asking for more or less elaboration of specific components, agents add information about possible obstacles. These conversations are more like problem solving or design than they are like simple recognition or recall. They comprise a complex and nuanced interaction.

One of the keys to understanding this interaction, we found, is seeing that *trust* between agent and client is a critical factor to the initiation and continuation of conversation between interlocutors.

Tailored

The agents have access to a few personal details about the clients that are relevant to providing guidance and advice. They use knowledge about the type of visual impairment the client has, their preferences about distance measures (e.g., meters, feet), and the task context (real world knowledge about domains) to provide help. When an agent is connected to a client, they can quickly scan profile information and assess where the client is and what they are trying to accomplish in order to adjust guidance and advice, tailoring it to the client's physical situation and preferences.

Agents told us that experienced clients have learned strategies to help agents help them. For example, the client knows about their home and neighborhood, their work environment, and places they regularly shop. They use their pre-existing knowledge of these places to quickly train up a new agent so that the pair will be able to more rapidly build common ground in future interactions.

Clients and agents need to trust one another in order for advisory interactions to succeed. They are inclined to assume trustworthiness, and then seek encouragement for this assumption as the interaction goes forward. When the agent demonstrates to the client that they took the trouble to identify a few relevant personal details, swift trust is strengthened [2].

Responsive

The agents we studied convey empathy and personal awareness to the clients. They specifically try to "read" the client's mood through speech, pauses, and reactions to guidance and advice that the agent offers through the course of the interaction. Thus, agents told us that if they noticed that a client seem disengaged, anxious, or "down", the agent focused on providing information required without conversational elaborations and side topics.

Being responsive to an interlocutor's non-verbal signalling is fundamental to successful conversation. But more specifically, it demonstrates concern for the interlocutor, and commitment to the conversation. Thus, beyond just facilitating successful interaction, it facilitates trust. A partner that conveys interest in and responsiveness to one's emotional state, is a partner one might trust.

Acknowledging

One of the most basic elements of conversation is acknowledging that a partner said something, and indicating what was understood in what they said. This includes content-neutral acknowledgements like "uh-huh" and "okay", but also explicit confirmation such as repeating a statement in the form of a question, or, most bluntly, asking directly whether the partner said what was heard.

Such acknowledgements are directly critical in that they curtail outright mistakes. But they also continuingly assure the partner that they are being heard, and that what they say matters enough to be verified. In the Aira interactions, this conveys that the client and the client's input is important. But for any conversational agent, acknowledgement conveys that the agent is actively processing and thinking critically. Who would trust an agent that was not?

Adaptive

Agent told us that clients regularly walk just a bit faster than the agents feel they can comfortably keep up with describ-

ing layout and scanning for obstacles. Agents report feeling they need to talk fast and look harder at the video display. Pushing the pace puts the client in more immediate control of the interaction, but makes the agent's job more challenging.

The agents try to keep up, but they also try to negotiate, sometimes asking the clients to slow down. This is affirming to the clients in two ways. First, it is a signal that they, the clients, are initiators in this interaction, that they can set the pace. But second, it can strengthen trust development by presenting a problem to the pair of actors that they must address together. It is indeed critically relevant for the client to wonder whether the agent can say something under time constraints that will help them (the client) perform: what can you say in 2 seconds that will help me make this point, explain this slide, answer this question, avoid this obstacle?

This is important since the agent has a important safety role, and needs the client's trust. Pushing the interaction (slightly) helps to assure both that the interaction is working.If it goes well, the client sees vividly that the agent is willing to work together, and is committed to providing good guidance and advice even when they need to work a little harder to do it.

Coordinative

The agents coordinated their advisory contributions closely with the client's behavior. They often precisely interleaved their guidance with the actions and reactions of the client.

Giving a presentation was a particularly interesting example of client/agent coordination. In this interaction, the person with visual impairment is presenting to a (mostly sighted) audience of other people. The agent provides an interleaved back-channel narration of audience reactions, and suggestions for adjustments to the presentation. The agent's objective is to provide guidance but to not disturb the flow of the presentation interaction. This back-channel guidance is facilitated by deliberate pausing and elongated word pronunciations on the part of the client/presenter, creating an intricate 3-way interaction in which, ideally, the audience experiences only the presenter.

Designing conversational agents

The possibility of conversational agents can benefit from understanding how human agents support clients through nothing more than conversation. The Aira conversational prosthetic is an example of this. It is a successful and popular service among people with visual impairment. One example of human-human interaction cannot dictate the design of every computational conversational agent, but such examples are an existing source of wisdom and experience that should be leveraged as part of a strategy to learn more and do better.

REFERENCES

[1] Sooyeon Lee, Madison Reddie, Chun-Hua Tsai, Jordan Beck, Mary Beth Rosson, and John M. Carroll. 2020. The Emerging Professional Practice of Remote Sighted Assistance for People with Visual Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

[2] Debra Meyerson, Karl E Weick, Roderick M Kramer, and others. 1996. Swift trust and temporary groups. *Trust in organizations: Frontiers of theory and research* 166 (1996), 195.

[3] Joseph Weizenbaum. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 9, 1 (1966), 36–45.