Software agents or intelligent personal assistants are increasingly prominent and ubiquitous, and already play a useful role in our everyday lives. Examples of such intelligent personal assistants include Microsoft’s Cortana (Cortana, 2015), Apple’s Siri (Siri, 2015), or Google’s Google Now (Google Now, 2015). These systems are able to process data, perform services, and even teach us. For anything other than very simple tasks, these agents are likely to need to engage in meaningful dialogue with their users, and understand their preferences and patterns of behavior over time. Agents can collaborate with users or perform tasks on users’ behalf autonomously, gradually becoming more effective as they learn the users’ interests, habits and preferences (Maes, 1994). Maes states that two main problems need to be solved when building these kinds of software agents: 1) Competence: decisions of when to help the user and what to help them with, and 2) Trust: how we can guarantee that the user will trust the agent in making those decisions?

Designing virtual personal assistants that effectively engage with users is not a new problem for HCI research. However, recent advances in speech recognition and natural language processing have made conversational systems practical in a broader number of settings, presenting new challenges for interaction research. While these systems may now recognize what an individual is saying, the lack of emotional intelligence they possess is marked. For example, an individual seeking advice about a distressing situation is unlikely to receive a very deep or empathetic response from a virtual assistant that exists today. However, research is beginning to address these kinds of situations.

Previous research has highlighted the many benefits of conversational systems if they can respond to non-verbal cues. For instance, they can encourage more engaging conversation between the human and agent (Cassell and Thorisson, 1999). They can also have better rapport (Gratch et al. 2007) and the ability to capture signs of bonding (Jacques et al., 2016) during the human-agent interaction. Emotionally expressive agents have also been found to be more trustworthy (Brave, Nass and Hutchinson, 2005; Bickmore and Cassell, 2001) and agents that match the user’s linguistic style have also been shown to be rated as more human-like, intelligent, trustworthy and engaging in their conversations with the user (Shamekhi, Czerwinski, Mark, Novotny, Bennett, 2016). In addition, when the conversational agent takes on a similar personality style to the human, the human considers the agent to be more engaging (Nass and Reeves, 2000). Research has shown that interacting with an attractive agent, even if that agent is less reliable than a neutral alternative, results in humans rating that agent as more likeable, trustworthy, accurate and intelligent (Yuksel, Collisson and Czerwinski, 2016). All of these findings point toward the need for a conversational agent to be designed as emotionally expressive, attractive and to match the user’s personality type and linguistic style. These characteristics will increase the likelihood of the user interacting with that agent and trusting it more.

Despite the apparent weakness of today's agents that lack emotional intelligence, it is still very challenging to design systems that address this. Effective conversationalists are adept at sensing, and adapting to, non-verbal cues. How to design a software system that can capture, make sense of, and respond to these subtle and often highly variable cues is still an open research question. Nuances in understanding conversational timing, pace, pitch, prosody, subtle facial and body gestures and even the kinds of words and loudness with which they are spoken can have major effects on a user's willingness to trust and interact with that agent.
Our research group has carried out a series of experiments looking at the impact of an emotional, conversational agent on users' subsequent interactions, and the users' own emotional states (both before and after the interaction). We will share design lessons that we have accumulated to date on these topics. One particular project is PocketSkills, which we intend to use as a future platform for exploring more of these issues. PocketSkills is a mobile phone application (and a website) that uses an emotional agent to converse with a human who would like to learn Dialectical Behavior Therapy, or DBT (Linehan et al., 1991). In our first user study with patients already in DBT therapy, we observed that a neutral to encouraging and even happy agent's conversational style was well accepted and trusted by this small set of users. Our next step will be to begin to apply machine learning techniques to data from a larger user population to determine which conversational style, emotional state and DBT skills content should be utilized by the agent.

References


Biographies:

Daniel McDuff is currently a Researcher at Microsoft Research in Redmond, WA, working on the design of affective computing systems. Daniel's research is at the intersection of emotion, computer vision, machine learning and HCl. He is interested in building hardware and software for human-sensing at
scale, and in technologies that improve well-being. Prior to working at MSR, Daniel was Director of Research at Affectiva in Cambridge, MA. Daniel received his PhD in CS in 2014 from MIT Media Lab, advised by Dr. Rosalind Picard.

Mary Czerwinski is a Principle Researcher and Research Manager at Microsoft Research in Redmond, WA, where she studies the impact of emotional, conversational systems on human behavior, in addition to many other areas of human-computer interaction. Mary received her PhD in 1988 from Indiana University in Cognitive Psychology.