A Future for Personal Assistants

Adam Cheyer
Viv Labs
San Jose, CA
adamcheyer@aol.com

Rana El Kaliouby
Affectiva
Waltham, MA
kaliouby@affectiva.com

Phil Cohen (moderator)
VoiceBox Technologies
Bellevue, WA
philipc@voicebox.com

Steve Whittaker (unconfirmed)
University of California
Santa Cruz
Swhittak@ucsc.edu

Eric Horvitz
Microsoft Research
Redmond, WA
horvitz@microsoft.com

Abstract

The purpose of this panel is to explore issues that will arise in building future *personal assistants (PAs),* especially for family use. In this regard, we will consider implications of being an "assistant" and those of being "personal." The target timeframe is 3-10 years out, so that near-term products will not be discussed. We will elaborate briefly on the kinds of communicative and inferential capabilities such PAs will need, and then examine their social and emotional capabilities. We will discuss pros and cons for their evolution and deployment. In this regard, we will discuss the kinds of support that could be provided by the HCI community in building personal assistant systems that are useful, delightful, functional, controllable, educational, ethical, and secure.

Author Keywords

Personal assistant, intelligent agent, multimodal dialogue, group communication, emotion recognition

ACM Classification Keywords

H.5.m. Information interfaces and presentation

Introduction

Scene: Sometime in the coming decade, at your house.

So you finally bring home your personal assistant that has been upgraded to "group capable." It knows your quirks, your fondest desires, your habits, your emotions, everything that a good personal assistant should. For instance, it knows that when your jaw is tensed, you are having a bad day and are going to chew someone out. It's now ready to meet your family, which it has learned about from your descriptions. However, it has not "met" them in any respectable fashion. Finally, it knows each family member has his/her own assistant.

> You arrive home, schedule a home meeting at which you turn on a screen. Your assistant "Mina" (you aren't so good with names) appears on the screen and proceeds to introduce itself to your family and their assistants. At some point, your teenager won't cooperate; your jaw tenses…
>
> What does your assistant do? As with the "Mad Libs" game, we get to write the ending to this scene. It could end in a delighted family learning how they can enjoy the new assistant in their midst. Or, it could end in dysfunction.
>
> How then do we get the right ending?

By now, many people have tried various "personal assistant" software applications, including those from the major telephone handset and computer manufacturers, as well as those offered by the various automobile companies. Every day new "capabilities" are touted in the press ("my assistant can handle this sentence …"). Such assistants are becoming more common for home and family use, and enthusiasm is building for their use in the so-called "Internet of Things." However, before we become giddy through hyperventilation, perhaps it's time we thought deeply as a community about the trajectory we are on.

The basic questions we need to ask are "what can these PAs do?" and "whose assistants are they?"

Assistance

First, we will distinguish between an "assistant" and an "agent". An assistant helps you get something done, including educating and amusing you, while we will say that an agent acts in your stead. For example, consider a real estate agent, who can list and show your house, and in exchange, has certain legal obligations and rights. Clearly, agents are a special case of assistants.

Assistant systems in early 2016 can engage in limited spoken language conversations, primarily to answer simple questions, but also to perform a small range of functions, such as to book reservations or provide directions to points of interest. The dialogues are mostly of the "slot filling" variety. The assistants make limited inference about users' intentions and plans, and generally do not go beyond providing a literal answer to a literal question , unless such responses have been preprogrammed.

A few PA systems allow alternative and/or simultaneous multimodal input, such as touch input, and gestures (facial and/or body). Previous CHI meetings have explored the topic of spoken/multimodal language and dialogue technologies, so this panel will examine broader issues that become relevant now that the basic communication technologies are (perhaps) beginning to become adequate.

Personal assistants as described above have so far been invoked in order to accomplish relatively atomic tasks at the time of utterance. They tend not yet to be able to develop their own plans for accomplishing a request, given standing orders, or given autonomy to accomplish tasks on their own. More generally, personal assistants don't yet provide much assistance. In fact, they don't yet DO much at all[1]. Although much work is underway to expand the "domains" in which they operate, such domains tend to be artificial constructions, making cross-domain discussion and action difficult. For the next decade, we may imagine and work toward a world in which the assistants are empowered as agents to act on our behalf in both an open physical world and in cyberspace, to negotiate with other agents, to make inquiries with people who

[1] With apologies to John Austin and the speech acts literature.

are can perform actions or are thought to know the answers to the questions at hand, etc. Approaches to these capabilities have been attempted previously, but have not been robust or scalable. Specifically, in order to build such a future generation of assistants, we will need them to:

- Recognize the users' intentions and plans
- Act to further those plans
- Collaborate with people and other assistants
- Engage them in (multimodal) conversations

In carrying on dialogues with people, such assistants would need to behave in fashion that people expect. For example, people should be able to request that the assistant perform an action at some future time, to which the assistant may reply with a commitment to do so, and then acts autonomously to fulfill that commitment.

Personalization

Many products are targeted at providing an assistant that has been personalized to a given user. In general, personalization is a complex process involving person identification via biometric data, attribution of personal preferences, representation/encoding of people's mental states, including dynamic beliefs/knowledge, goals, plans, intentions, obligations, etc. Personalization is enabled by machine learning technologies, including learning by observation, by being told, by user demonstration, by (self) experimentation, etc. The whole subfield of User Modeling is relevant here, but is too large a topic to discuss in this panel.

Personalization interacts with an assistant's persistence in subtle ways. The first versions of (simple) assistants on the market have targeted mobile devices and automobiles, usually with cloud-based software. There seems to be a trend, favorable to the vendors, that there should be one assistant who acts for the user in every environment. When the user is changing environments, say from walking with a phone to driving in a car, the same assistant would thus be available in the second (auto) environment and the conversation should continue from where it left off. One way to signal that it's the same assistant would be to preserve its voice, personality, and appearance, and especially the prior discourse context. Yet if the user selects a different form factor, perhaps one with a screen or a robot, should the same assistant be available now embodied as an avatar or robot? How would it be able to take advantage of the new input and output modalities?

Now, imagine a situation in which the user leaves the car and enters the home. This poses many unanswered questions, such as how a personal assistant should act in a family setting? Should the assistant have *group intelligence*? That is,

- Should there be a single assistant per user, per family, per group, per institution? Or a personal assistant along with a family assistant that can gather information from the personal one?
- Should an assistant enter into either single or multi-party dialogues with members of the group or with their assistants?
- Can the assistant be trusted to act as its user's "agent", acting on behalf of the user when interacting with other assistants or people?
- Will the interaction styles of a family be reflected in those of the family's assistant(s)?
- How should assistants handle conflicting goals and requests of group members? (Imagine a teenager who wants his/her own assistant not to share information with other family members).
- Given the diversity of opinions in families, how will their assistants built by different vendors interact? For example, would they interact through natural language, enabling humans to

monitor their conversations, or interact via specialized computer languages? Dare we ask - will there need to be standards?

One topic that will certainly affect family-based interaction with assistants is that those human-human interactions are often emotionally laden. The area of Affective Computing is starting to blossom, and promises to provide attributions of users' emotional states through observation of various signals, such as from face, voice, stance, etc. One can imagine many important but so far nascent applications for emotion recognition within the personal assistant context. This leads us to ask:

- How accurate is emotion recognition now, and how accurate does it need to be to take useful action? What can be expected of this technology over the next decade?
- What emotional states are important to track?
- What sensing modalities and their combination are optimal?
- Can the user be characterized as being in multiple emotional states to varying degrees simultaneously? How can different modes signal conflicting emotions simultaneously (e.g., irony)?
- How should an assistant react to the emotional state(s) of the user? E.g., if the user is recognized as being depressed, what should the assistant do? Should the assistant delay action until the user is in a happier state? What if the user appears to be ready to take an action that would be injurious to him/herself or someone else?

Privacy, Security, Ethics

The general public is beginning to understand the tradeoffs with current assistant technologies in terms of providing personal data to vendors and in exchange for receiving better performing software. But in doing so, the user exposes data that can potentially be used in unanticipated ways. Security and privacy have become an increasing concern for children's interactions with assistant technology. Although sufficient performance may be available in the next decade to support local processing, the assistants may still need to interact via the internet with others. How do we design the appropriate levels of privacy and security to enable the benefits of personal assistant technologies? Do we need a codified ethics for assistants (similar to Asimov's laws of robotics), which will guide what an assistant system *should*/*should not* do? One group with which to begin this important discussion is the HCI community.

Role of HCI Research

HCI research has many important roles to play in helping to forge useful, usable, delightful, ethical, entertaining and educational personal assistants. Relevant research includes (but obviously is not limited to) scientific studies of language and cognition, investigations of social situations that include software assistants, and support for individuals with disabilities. One key outcome of this panel would be to stimulate a discussion of how relevant HCI research can play a more direct role in guiding the next generation of these technologies?

Role of Panelists

The panelists (if there are 4 of them besides the moderator) will be given 10 minutes for a position statement. If there are 5 panelists, opening position statement time will be 7 minutes apiece (assuming transition time). We expect there will be considerable exchange among the panelists and especially with the audience. Thus, we are leaving half the available time for audience questions.

Panelists

Adam Cheyer
Co-Founder and VP Engineering
Viv Labs

Bio: Adam Cheyer is co-founder and VP Engineering of Viv Labs, a startup whose goal is to simplify the world by providing an intelligent interface to everything. Previously, Adam was co-founder and VP Engineering at Siri, Inc., When Siri was acquired by Apple in 2010, he became a Director of Engineering in the iPhone/iOS group. As a startup, Siri won the Innovative Web Technologies award at SXSW, and was chosen a Top Ten Emerging Technology by MIT's Technology Review; Apple's version of Siri was presented "Best Technical Achievement" at the 2011 Crunchies Awards, and is now available on hundreds of millions of devices.

Adam is also a Founding Member and Advisor to Change.org (130M+ people taking action, victories every day), and a co-founder of Sentient.ai (solving the world's hardest problems through massively-scaled machine learning). As a researcher, Adam authored 60 publications and 23 issued patents. At SRI International, he was Chief Architect of CALO, one of DARPA's largest AI projects. Adam graduated with highest honors from Brandeis University and received the "Outstanding Masters Student" from UCLA's School of Engineering.

Phil Cohen (moderator)
VP Advanced Technology
VoiceBox Technologies

Bio: Phil Cohen has long been engaged in research in the areas of natural user interfaces, multimodal interaction, intelligent agents and multiagent systems, and human-computer dialogue. At VoiceBox Technologies, he is charged with developing next generation multimodal dialogue systems, for mobile, automotive, and home use. At Adapx Inc., which he co-founded, he co-led the development of the Sketch-Thru-Plan system for multimodal course-of-action creation. He also led projects on digital paper applications for maintenance data capture, for field medical data collection and for traumatic brain injury assessment. At the Oregon Graduate Institute, he co-led research with Dr. Sharon Oviatt on multimodal interaction. At SRI International, he engaged in research on the foundations of intelligent agents and multiagent systems, as well as research on multimodal interaction. Among the best decisions he made there was to hire Adam Cheyer to work on the Open Agent Architecture, which ultimately led to the development of Siri. Cohen is a Fellow of the American Association for Artificial Intelligence, and has been President of the Association for Computational Linguistics. He has more than 150 journal and conference publications. Cohen is the recipient (with Prof. Hector Levesque) of an inaugural Influential Paper award by the International Foundation for Autonomous Agents and Multi-Agent Systems for his research on the theory of intention.

Eric Horvitz

Technical Fellow at Microsoft
Managing Director, Microsoft Research, Redmond

Bio: Eric Horvitz has been pursuing research in artificial intelligence and multimodal interaction, focusing on principles and applications of machine perception, inference, and decision making. He has been elected fellow of the National Academy of Engineering NAE), ACM, AAAI, and of the American Academy of Arts and Sciences. He was inducted into the CHI Academy for contributions including "principles of mixed-initiative interaction for interleaving automated services with user actions, methods for predicting intentions and goals of users, and advances with spoken dialog systems." In 2015, he was awarded the AAAI Feigenbaum Prize "for contributions in computational models of perception, reflection and action, and their applications., and the 2015 Sustained Achievement Award from the International Conference on Multimodal Interaction. He received his Ph.D. and M.D. degrees from Stanford University. More information can be found at http://research.microsoft.com/~horvitz/.

Rana El Kaliouby

Co-Founder and Chief Scientific Officer
Affectiva

Bio: Rana el Kaliouby, PhD, is Chief Strategy and Science Officer and Co-founder of Affectiva, the global leader in emotion sensing and analytics. She invented the company's award winning, automated emotion-sensing and analytics technology, used by over 1,400 brands and one third of the Fortune Global 100 today. Prior to that, as a research scientist at MIT Media Lab, Rana spearheaded the applications of emotion technology in a variety of fields, including mental health and autism research. Her work has appeared in numerous publications including The New Yorker, Wired, Forbes, Fast Company, The Wall Street Journal, The New York Times, CNN, CBS, TIME Magazine, Fortune and Reddit. A TED speaker, she was recognized by Entrepreneur as one of the "7 Most Powerful Women To Watch In 2014", inducted into the "Women in Engineering" Hall of Fame, a recipient of the 2012 Technology Review's "Top 35 Innovators Under 35" Award, listed on Ad Age's "40 under 40" and recipient of Smithsonian magazine's 2015 American Ingenuity Award for Technology. Rana holds a BSc and MSc in computer science from the American University in Cairo and a PhD from the computer laboratory, University of Cambridge.

Steve Whittaker (invited, unconfirmed)

Professor
Human-Computer Interaction
University of California, Santa Cruz

Bio: Technology is transforming our everyday lives, how we think and interact. I work at the intersection of Psychology and Computer Science. I study how technology is affecting fundamental aspects of our everyday lives, and use insights from Cognitive and Social Science to design new digital tools to support effective focus, memory, collaboration and to help manage personal information. My past research was funded by the EU, NSF, EPSRC (UK), Google and Microsoft. I currently have an NSF grant to research Technology Mediation for Emotion Regulation, and a Google grant looking at collaborative file sharing. I am Editor of Human Computer Interaction 1 of 2 top HCI journals. Recently I had the huge honour of being awarded a Lifetime Research Achievement Award from SIGCHI, the society of Human Computer Interaction professionals. I am also a Fellow of the Association of Computational Machinery (ACM).