
“OK Google”: Building a Conversational Search Assistant

Nikhil Sharma

Google, Inc.
1600 Amphitheater Parkway.
Mountain View, CA 94043 USA
nikhilsh@google.com

Abstract

Recent advancements in speech recognition and aspirations of making information easily accessible to more people in more situations have led to a renewed focus on voice/speech use. With this renewed focus we have also come up against challenges to using voice effectively for HCI. I present four important challenges that I hope to work to overcome with other scholars engaged in making voice/speech use the future of HCI.

Author Keywords

Copyright 2014 held by Owner/Author. Publication Rights Licensed to ACM

Voice; search; conversation; speech; assistant; actions.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

To organize the world’s information and make it universally accessible and useful, we need to employ multiple modes of interaction. There are many situations where speech is the most appropriate mode for making information accessible and useful. Recent advances in speech technologies like support for different languages and touch-free hot wording have led to an increased focus on voice as an important interaction mode.

In this paper I will talk about my background, some speech focused projects that I have worked on and some challenges that I have come across. My hope is to have some of these challenges considered in the workshop.

Background

My doctoral studies and my multiple HCI publications focused on how people search for and make sense of information. Since I have always been very interested in how HCI can help people get to the right information,

working on Search was a great fit for me after my PhD. I am currently the research-lead for voice search at Google. Over the last year or so I have worked on establishing a research vision for voice use in search. This research vision is needed to support large-scale voice efforts that are currently underway. I have conducted more than 15 studies on voice interaction using a variety of methods from controlled experiments to surveys.

Building a Conversational Search Assistant

In Google, as well as industry in general, there is a renewed focus on voice. One initiative has been to make voice interaction available across devices. This includes making devices that interact primarily through voice (e.g. Google Glass, in-car systems), making voice central in devices that currently utilize voice (e.g. smartphones, tablets) and introducing voice to devices where voice use was not present before (e.g. desktops).

Another initiative has been to provide assistance to the using voice wherever possible: for search [how tall is the Eiffel Tower] as well as actions [Call mom on her mobile phone]. This renewed focus has highlighted a lot of challenges that have generated the need of research. I will briefly write about four of these challenges.

1. Affordance for a Natural Interaction

Most people are able to communicate with each other using voice and natural language. A simplistic view of voice interface use would suggest that people could also easily interact with systems using natural language. However, due to years of keyword based system interaction and because of other reasons like lack of trust in the system, we may have trained people

to not use natural language when they interact with systems. People don't easily transfer natural language usage to voice-based computer systems. This presents a challenge for deployment and use of voice based systems. How can we create systems that have the right affordance for users to pick up and use them? How can we encourage people to interact with the system using natural language rather than keywords?

2. Being Natural vs. Being Explicit

Another challenge for product design is that usability heuristics are often seemingly at odds with each other. If not unique to voice, this dilemma is quite central in voice-based systems. For example, we might have a goal for a voice-based system to be natural, human-like and conversational. This might mean using pronouns instead of proper nouns, not repeating what was recognized and to be concise. Humans dislike repeated verbal instructions [1]. However, in a voice only system conveying system state becomes very important so that users can recognize and recover from errors. Imagine asking your phone to call a friend with a difficult to pronounce name. If the phone responded by just calling the person, you may well be concerned that the phone may have misinterpreted you and called the wrong person. Users need explicit spoken feedback and system-state information when they are in eyes-free modes. Deciding when to be natural/human vs. when to be explicit is a great challenge.

3. Enabling Correction

Voice based systems can offer speed, efficiency and convenience on small devices where touch interactions are laborious. Errors of recognition often wipe away the speed and convenience advantage of voice interaction. One important usability heuristics [2] is that interfaces

should allow users to identify and recover from errors. This means that an ability to easily correct mistakes is very important for voice-based systems to utilize their full potential. One particular challenge for correction is having the right affordance so people can fix misrecognition through voice input.

4. Choosing Personality

Since voice use developed in a social context, its use even by a system or machine often results in attribution of personality [1]. Voice based systems, whether they choose to or not, will usually evoke a personality in the mind of the user. The dilemma then becomes what kind of personality should a system assume. A system can choose to have a fun/likeable personality or an efficient and capable personality. Choosing a personality will guide how the system responds and also what users will expect from it. However, we interact with other humans of varying personality for different purposes. For example, we may want our surgeon to be competent but our bartender to be fun. Similarly, users may desire different personality traits from a voice based system depending on their activity and frame of mind. The challenge thus becomes not only in choosing the right

personality for the system but also being able to adapt that personality depending on the user's situation.

Workshop

The four challenges do not compromise a comprehensive list. These and many other challenges have created research needs to keep a community of scholars busy for a long time. I am interested in hearing what other challenges have been encountered? What methods have been used to study them? What solutions have been tried and rejected? What has worked? More than that, I would like the opportunity to engage with scholars from both academia and industry who share the common belief that voice/speech offers a challenging but exciting opportunity for human-computer interaction.

References

- [1] Nass, Clifford Ivar, and Scott Brave. *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge: MIT press, 2005.
- [2] Nielsen, Jakob. "Ten usability heuristics." (2005): 2008.