'Siri Thinks I Have Two Wives' & Other Embarrassing Voice Interactions

Nikhil Sharma

Google, Inc.

1600 Amphitheater Pkway.

Mountain View, CA 94043 USA

nikhilsh@google.com

Abstract

Voice use is gaining prominence. However, people feel awkward using voice user interactions (UIs) around others. This is because voice UIs can be particularly embarrassing. Based on two years of research on voice UIs, I suggest some reasons why voice UIs can be particularly embarrassing. I also list research questions that can help guide development of voice UIs. I hope the workshop will provide an interesting forum to discuss challenges as well as opportunities for reducing embarrassment while using voice UIs.

Author Keywords

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.

However, there are many barriers to the adoption of voice UIs. One of them is that they have the potential to be embarrassing. In this paper I will talk about my background, some voice-focused projects that I have worked on that have examined barriers and challenges to voice UI adoption. My hope is to have some of these challenges considered in the workshop, especially those related to embarrassing interactions.

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 two years 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 30 studies on voice interaction using a variety of methods from controlled experiments to surveys. While this position paper does not include detailed findings from those studies, I am looking forward to sharing and gleaning insights at the workshop.

Voice & Embarrassment

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, Android Wear watches, 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].

Along with great possibilities for interactions, voice also brings potential for embarrassing interactions.

Figure 1. Siri thinks I have two wives. Source: Sean Herber [1].

Imagine the situation in Figure 1 happening when the user is with other friends. The situation can be comedic or disastrous depending on whether the "wife" is present.

Siri is by no means the only voice UI that creates embarrassing situations. In my studies, when I have asked users what some of the barriers to using voice UI are, 'being socially awkward' is one of the most frequently mentioned.

This makes one wonder, what makes voice interactions socially awkward and embarrassing. I have postulated three reasons why this might be the case:

First, while voice recognition is improving it is imperfect; voice interactions also involve significant artificial intelligence that is not trivial. Thus the possibilities for errors are common.

Second, voice interactions have a greater potential to be overheard. The progression from time-share computers to personal computers to smartphones and tablets has made computing devices very personal. People can use touch UI to interact with their devices and the smaller screens of smartphone offer more privacy. Voice UIs run counter to this trend.

Thirdly, voice as an interaction method co-evolved with our social interactions. This makes voice an inherently social human-to-human interaction medium [1]. This could be a reason why people feel awkward using voice UIs to interact with a device when other people are around.

Overcoming Embarrassment
All this raises interesting research questions about voice interactions and embarrassment:

Firstly, can voice interaction technology become sufficiently error free to avoid embarrassing situations? More importantly, since no UI is error free, how can voice UIs be designed that errors do not embarrassing our users?

Secondly, will users feel comfortable using voice UI in presence of others? On one hand voice interaction with a machine could be an inherently awkward situation, on the other hand people have adopted many technologies that once seemed socially awkward. Talking on a mobile phone and taking 'selfies' in public are some examples. If a critical mass of users start using voice UIs in public, will it become the norm to use voice UIs in public? Initial data suggests that teens may be more comfortable using voice than adults [2].

These questions do not compromise a comprehensive list. These and many other questions will keep a community of scholars busy for a long time. I am interested in hearing what other UIs have faced and overcome embarrassing interactions? 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 embarrassment should not be the cause for stopping, but for learning and forging ahead.

Reference

[1]   Nass, Clifford Ivar, and Scott Brave. Wired for speech: How voice activates and advances the human-computer relationship. Cambridge: MIT press, 2005.

[2]   Sean Herber. Great! Now Siri thinks I have two wives. Twitter post. https://twitter.com/bigzaphod/status/125694312305537024

Retrieved: Retrieved: January 19th, 2015

[3]   OMG! Mobile voice survey reveals teens love to talk. Google Official Blog. http://googleblog.blogspot.com/2014/10/omg-mobile-voice-survey-reveals-teens.html

Retrieved: January 19th, 2015