

# Voice activated intelligent personal assistants: challenges and opportunities

**Maria Gini**

**Department of Computer Science and Engineering  
University of Minnesota**

# Motivations and scope

- This talk is about opportunities and challenges, not about specific research results. I was inspired by the workshop statement “*This year's workshop will be centered on application domains.*”
- This talk is to try to convince you that there are low hanging fruits for this community, which can make a big impact on the lives of people.
- I will outline the opportunities and challenges, and open a conversation on potential solutions.

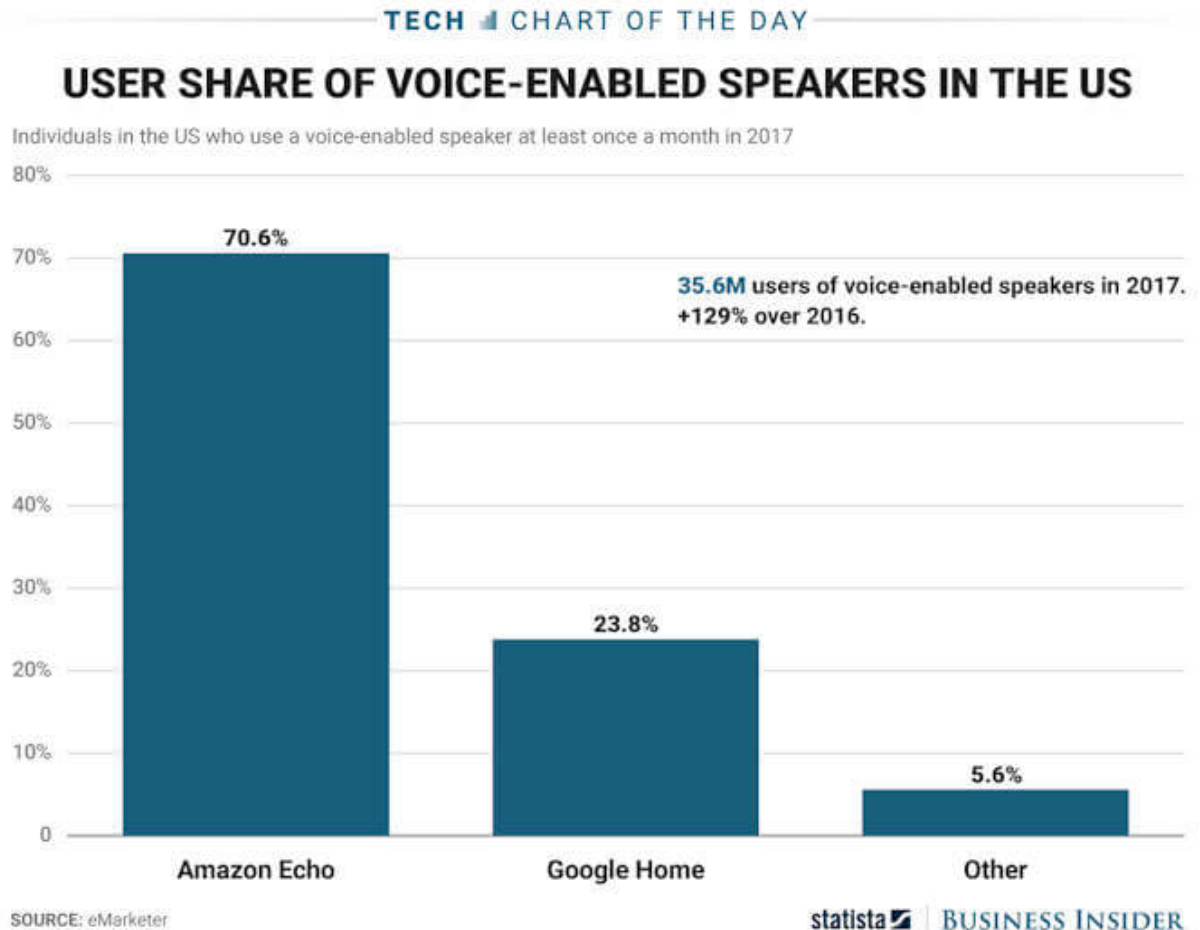
# Opportunity: Voice Activated Personal Assistants

- A large variety of voice activated personal assistants has become available in the last few years
- Around 700 million people are using some form of AI personal assistant (35,000 users of Alexa in 2017).
- The market is expected to grow to almost 2 billion by 2021.
- There are multiple choices: Siri, Google Assistant, Amazon Alexa, Microsoft Cortana. Many more to come!



# Voice-enabled speakers

Voice enabled speakers are powered by Alexa (for the Amazon Echo) and by Google Assistant (for Google Home). The voice based personal assistants are at the core of their functionality.



# Mystery.....

Is Alexa about to get a new voice? That's the big question behind Amazon's newest Super Bowl ad.

The answer will come "2.4.18"

Mystery solved: "We do alter our Alexa advertisements ... to minimize Echo devices falsely responding in customer's homes." A 2014 patent titled "Audible command filtering" details the techniques that keep Alexa from waking up in your home. "The patent broadly describes two techniques. The first calls for transmitting a snippet of a commercial to Echo devices before it airs. Then the Echo can compare live commands to the acoustic fingerprint of the snippet to determine whether the commands are authentic. The second tactic describes how a commercial itself could transmit an inaudible acoustic signal to tell Alexa to ignore its wake word."

# Why it this an opportunity?

- The current systems provide multiple skills (Alexa) and actions (Google). Their numbers continue to grow.
- Many software tools from Amazon and Google are available to add new skills/actions.
- Many devices have been sold.
- The devices are very affordable.

# Why voice activated assistants?

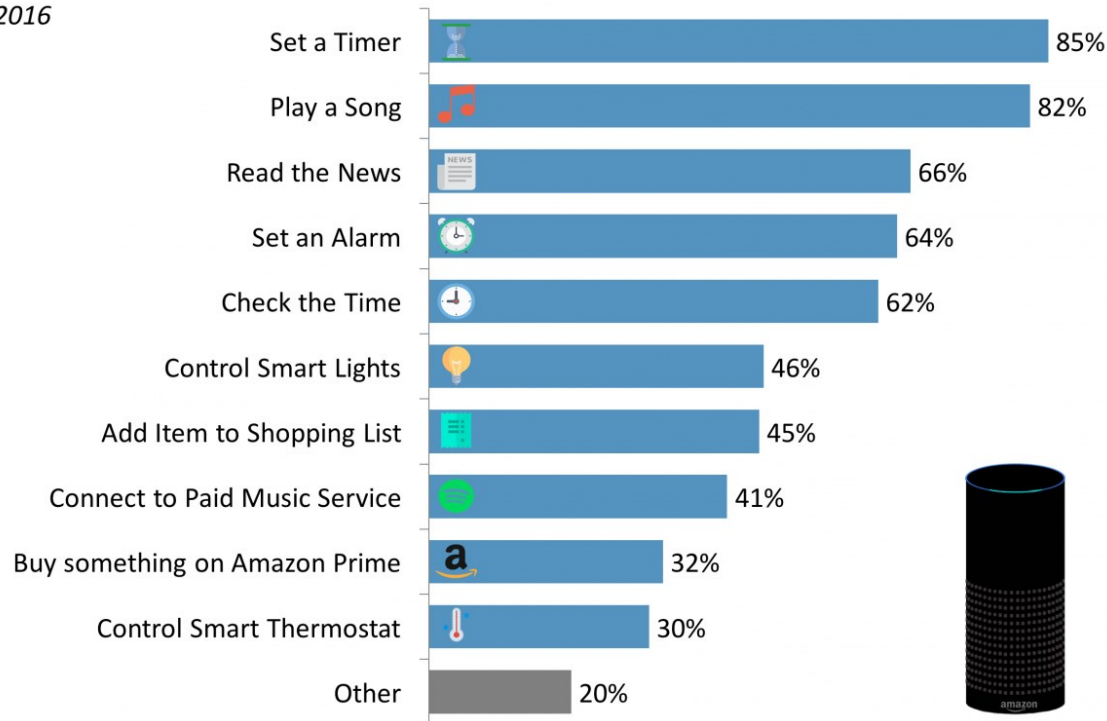
- The practical lack of keyboard and small screen of many devices will make voice the method of choice for communication.
- The main limitation is the reluctance of people to speak up when not at home. More than 60% of the users speaks up to their phone at home, but much fewer in public spaces.
- Voice enables users with different impairments to interact:
  - ▣ Hard-of-hearing users can connect the device to speakers or headsets.
  - ▣ Low vision users do not have to try and read input prompts or calendar data.
  - ▣ Users with motor control issues do not have to type or write any information.

# Is it really an opportunity?

## What do users do with these devices?

### Skills Echo Owners Have Used At Least Once

2016



Source: Experian Information Solutions

BI INTELLIGENCE



# Who are the users?

- Who are the users? The core users are millennials, baby boomers don't adopt them (>90% of baby boomers do not use them).
- Elderly tend to interact with intelligent systems in a manner much more consistent with the way they would interact with a human.

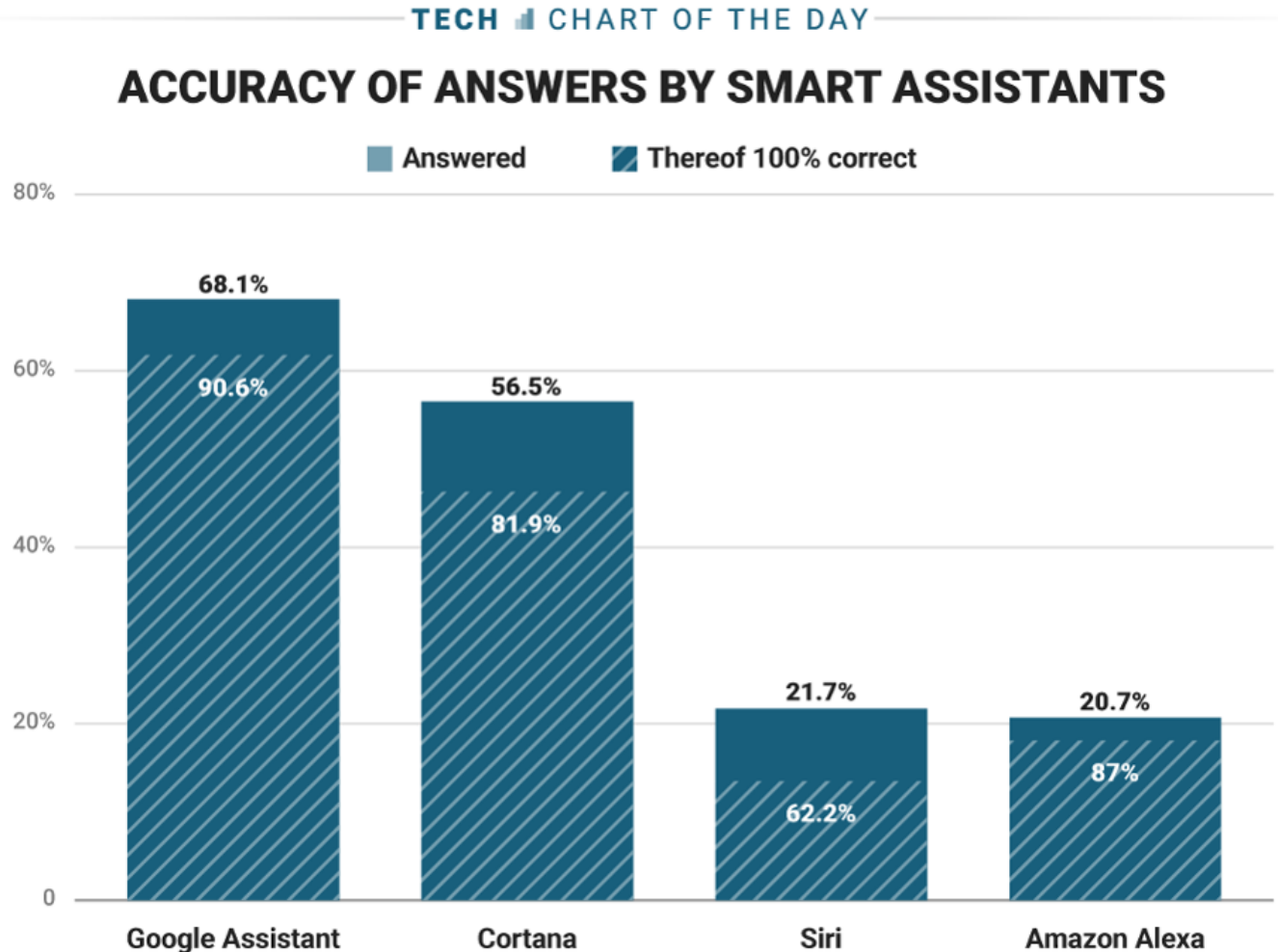
K. Georgila et al . A Fully Annotated Corpus for Studying the Effect of Cognitive Ageing on Users ' Interactions with Spoken Dialogue Systems. Proc. 6th Int'l Conf. on Language Resources and Evaluation (LREC), (1):938–944, 2008.

S. Moller, F. Gödde, and M. K. Wolters. A Corpus Analysis of Spoken Smart-Home Interactions with Older Users. LREC, pages 735– 740, 2008.

<https://techcrunch.com/2017/05/08/amazon-to-control-70-percent-of-the-voice-controlled-speaker-market-this-year/>

# Is accuracy in the answers important?

Accuracy of answers (as a percentage of correctness of the questions answered) does not seem to bother users



SOURCE: Stone Temple Based on a sample of 5,000 questions about factual knowledge

statista | BUSINESS INSIDER

# What are the user needs?

- Users of technology of all age groups have been shown to have at least a mild preference for speech interaction, either in a unimodal or multimodal system.
- Why many do not use them?
  - ▣ Poor voice interaction, inability to understand speech
  - ▣ Limited dialog management, inability to engage in longer conversation
  - ▣ No context awareness
  - ▣ No ability to adapt to users

P. Milhorat et al. Building the next generation of personal digital assistants. 1st Int'l Conf. on Advanced Technologies for Signal and Image Processing (ATSIP), pages 458–463, 2014.

Luger, Ewa, and Abigail Sellen. "Like having a really bad PA: the gulf between user expectation and experience of conversational agents." Proc. 2016 CHI Conf. on Human Factors in Computing Systems. ACM, 2016.

# Could other types of users use them?

For instance:

- Elderly people
- Blind/low vision people
- People with limited motor control
- People with mild cognitive impairment (MCI), for instance amnesic MCI where the memory loss is the main symptom

# Elderly people

Elderly users in general tend to:

- Interact with systems in a manner much more consistent with human-human interactions
- Use a richer, more variable vocabulary and social interaction words.
- Take more turns in dialog.

Speech specific

- Have a greater incidence of verbal pauses.
- Have a greater incidence of backtracking.

# People with sensory or motor impairment

Conservative estimate by NIH's National Eye Institute: 3.5 to 5 million Americans have **impaired vision** (letter acuity less than 20/60.) The World Health Organization (WHO) estimates there are 285 million people worldwide with vision impairment.

The Centers for Disease Control and Prevention report that 14% of US adults suffer from seriously debilitating **mobility impairments**.

NIH's National Institute on Deafness and Communication Disorders (NIDCD): 17% of the adult US population (around 36 million people) report some form of **hearing loss**, 50% of those aged 75 or older. Internationally, WHO estimates that 360 million people have "disabling hearing loss" (defined as a loss greater than 40 dB in the better of the two ears).

# Memory Loss

Mild cognitive impairment (MCI) encompasses a variety of symptoms and is often one end of a spectrum of cognitive issues that includes more severe disorders such as Alzheimer's disease and dementia. It is estimated that 15--20% of people age 65 or older have some form of MCI, and a significant number of those affected by MCI will go on to develop the more severe disorders

# Cognitive Resources and Coping

- Neurological disorders, such as MCI ,are often accompanied by increasingly limited cognitive resources due to chemical and physical changes in the brain (Bäckman et al. 2000). Cognitive resources affect capability to respond to and engage with many tasks, from routine daily activities to emotional labor to novel situations.
- Even in a healthy brain, the amount of cognitive resources available is finite, and so the more resources are devoted to routine daily tasks, the less are available for other processes. This is especially significant in MCI patients as memory loss issues are often comorbid with anxiety and depression, and in fact unmanaged anxiety has been shown to be a predictor for increased and more rapid cognitive decline (Beaudreau and O'hara 2008).
- Technological support for routine tasks can help free up cognitive resources for higher-level tasks, such as engaging in cognitive rehabilitation and the emotional labor of coping with anxiety, depression, and memory loss itself, that are vital to successfully treating, managing, or even slowing cognitive decline (Lindenberger et al. 2008) (Beaudreau and O'hara 2008).



# How to support users with dementia?

Meiland, F. et al. (2017) defined three areas where technologies can support persons with dementia:

- 1. managing their everyday life across the disease journey;
- 2. engaging in meaningful and pleasurable activities such as cognitive stimulation and physical activities, as well as improving social participation, contact, and support;
- 3. supporting professional organizations and systems within dementia health and social care, such as behavior monitoring, shared decision making, and Global Positioning System (GPS) tracking systems.

Meiland, F. et al. 2017. Technologies to support community dwelling persons with dementia: a position paper on issues regarding development, usability, effectiveness and cost-effectiveness, deployment, and ethics. *JMIR Rehabilitation and Assistive Technologies* 4(1).

# Opportunity: improve these devices

Dialog management falls short for some key demographics.

What can be done to improve it?

- ▣ Managing longer dialog histories.
- ▣ Improving context awareness, through dialog or additional sensor data.
- ▣ Dynamic, real-time adaptation to user behavior.
- ▣ Better handling of task hierarchy
- ▣ Learn user preferences

**Designing systems with specific user groups in mind** may lead to more robust, natural spoken dialog systems.

# ELIZA

Hello. How are you feeling today?

>I am happy

Very interesting.

>Is that all you can say?

Please consider whether you can answer your own question.

>Can you tell me about the weather?

Why do you ask if I can tell you about the weather?

>I think you should know it

Do you doubt me should know it?

>You do not respond

We should be discussing you, not me.

Weizenbaum, J. (1966) ELIZA - A computer program for the study of natural language communication between man and machine. CACM 9(1) 36-45.

# Our Initial Approach

Develop a complete calendar assistant controlled via voice commands to support people with mild cognitive impairment. The system will have:

- ▣ Agent-initiated conversational interaction
- ▣ Smart, context-driven default phrases
- ▣ Agent-initiated reminders
- ▣ Modeling of user preferences for personalized interactions
- ▣ The design will be user-driven

## System

- ▣ Amazon Echo device platform with Alexa voice services
- ▣ Syncs to cloud services for access outside the home
- ▣ Application and data hosted on Amazon web services (AWS Lambda)
- ▣ Natural language interaction through Dialogflow (API.AI) chatbot

# Why focus on the calendar?

- We have focused our efforts in specializing the personal assistant to handle adding/removing/changing events in the calendar.
- Effective time management is an important tool in coping with memory loss, though time management applications designed with this population in mind are still lacking (Meiland et al. 2017).
- Many people with MCI struggle with keeping track of time, and time and scheduling issues often cause stress among both patients and caregivers (Hawkey et al. 2005).
- An application capable of answering scheduling questions, no matter how repetitive, and keeping track of day-to-day activities could therefore have a significant impact on both patient and caregiver quality of life.

# Current state

- Tested features on both an Amazon Echo Dot and Google Home.
- The Google Home application uses Pull-String, a basic chatbot service, to interpret user speech and respond to prompts.
- Main goal is to evaluate the feasibility of storing, querying, and communicating calendar data that has not been synced to a calendar service, and then syncing to a calendar service independently of user input.
- All calendar actions are performed on local calendar data, which can then be pushed to a calendar service as a workaround to gaps in voice command support for both the Google Home and Alexa devices. Independent data representation and syncing has proven to be a challenge. We are still working on it.
- Libby Ferland, Ziwei Li, Shridhar Sukhani , Joan Zheng, Luyang Zhao, Maria Gini, Assistive AI for Coping with Memory Loss, Workshop on Health Intelligence at AAIL 2018.

# Commands supported by our prototype

Sample actions and commands supported by our prototype, matched with identified feasibility requirements. Actions in bold are not natively supported by either Alexa or Google home.

Command	Natively supported	Requirements tested
"Add an event to my calendar."	Yes	Data storage/queries
<b>"Remove lunch tomorrow from my calendar."</b>	<b>No*</b>	<b>Data storage/queries; account access; feature gap</b>
<b>"Change the appointment today at 11 am to 10 am."</b>	<b>No</b>	<b>Data storage/queries; account access; feature gap</b>

\* The ability to remove calendar entries has been added recently.

# Technical Issues

- **No native calendar support.** Alexa and Google home do not support voice modification of calendar events. We created a local representation of calendar data and then sync data at regular intervals. This means that calendar data may not be available outside of the home immediately. Storing and syncing calendar data from voice interactions has proven a significant challenge in and of itself.
- **Device storage and web hosting.** Data cannot be stored locally, so data that needs to be stored must go through a hosting service. Voice commands must also go through a hosting service.
- **Privacy.** To respond to voice commands, devices must be listening at all times. All user commands are recorded and stored on external servers. These voice clips include seconds of conversation before a wake phrase is detected, and it is not transparent how users can access or delete these clips. This may cause issues due to the nature of the user population – namely, the possibility of sensitive (eg., health) data being stored externally.



# Potential Added Value

Analysis of spoken utterances, either spontaneous or as part of standardized tests, may be effective as a diagnostic tool for early Alzheimer's disease or mild cognitive impairment .

Speech characteristics tend to become more pronounced as these diseases progress, so speech analysis may be a valuable tool in monitoring and developing treatment plans for memory disorders common in elderly populations.

Unfortunately, currently the Alexa APIs do not provide access to the speech, only to its transcription.

A. König et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease.

*Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(1):112–124, 2015.

# Discussion

- What can each of you contribute to make voice assisted personal assistants useful to a larger segment of the population?

# Conclusions

- Voice activated personal assistants offer a great potential for new uses.
- There are still many limitations in the APIs that make it difficult to extend their functionalities and make them appropriate for new types of users.
- Better dialog ability is needed to make them more widely usable. This requires context awareness, intent recognition, learning user preferences, richer language recognition.

# Thanks for your attention!

---

For more information go to  
<http://www.cs.umn.edu/~gini>  
or email to [gini@cs.umn.edu](mailto:gini@cs.umn.edu)