Spoken Dialog Design

JBS: Mobile Voice Applications
Lecture 1, Morning
June 7, 2018

+ Course Overview

- Syllabus, schedule, assignments, papers available at
  - https://sites.google.com/a/brandeis.edu/jbs-2018-cosi/courses/cs115aj
- Assignments:
  - Quizzes on vocabulary and readings
  - Occasional "Blog posts" on speech, apps, and other thoughts
    - Counts as class participation
  - Programming and presentation assignments

+ Project Timeline

Week 1	Come up with ideas Get to know your classmates Team building
Week 2	Coalesce into 5 groups, each with a project idea Broaden ideas and narrow scope
Week 3	Begin design
Week 4	Produce a “vision” clip Identify your “MVP”
Week 5	Select components Produce architecture diagram
Weeks 6-8	Present your MVP and your plan to build it BUILD!
Week 9	Final touches Showcase Submit final group evaluation

+ Tools we'll use

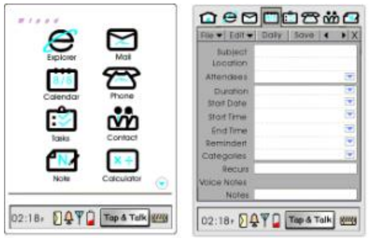
- Latte: Submitting quizzes and assignments
- Slack: Forum for discussion
- Googledocs: Sharing documents
- Github: Storing and sharing programs
- Lots more specific to speech and NLP
  - Dialogflow
  - AlexaTools
  - NIST scoring software

+ Broad goals

- “Experiencial Learning”
  - Learning from doing
  - Learning from each other
  - Learning from the many sources of information available
- Content goals
  - Understand the state of the art in
    - Speech recognition
    - Natural language processing for Dialog
    - Dialog management
  - Understand the process of designing a speech application
    - Use cases, scenarios, user profiles
    - Become operating system and platform agnostic
  - Develop proficiency in some of the tools available
  - Learn how to figure out new tools

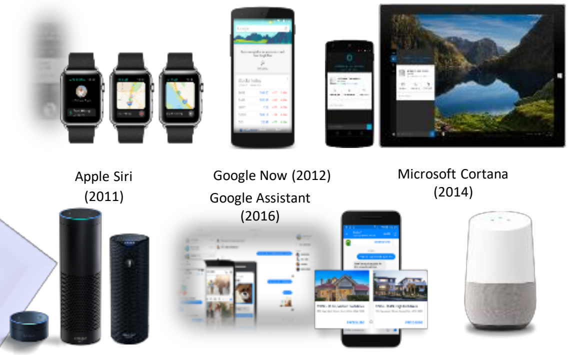
+ Brief History of Dialog Systems

Multi-modal systems
e.g., Microsoft MiPad, Pocket PC



TV Voice Search
e.g., Bing on Xbox

Virtual Personal Assistants



Apple Siri (2011)

Google Now (2012)
Google Assistant (2016)

Microsoft Cortana (2014)

Amazon Alexa/Echo (2014)

Facebook M & Bot (2015)

Google Home (2016)

Task-specific argument extraction
(e.g., Nuance, SpeechWorks)
User: "I want to fly from Boston to New York next week."

Early 20

Siri

2017

IBM WATSON

Early 1990s

Intent Determination
(Nuance's Emily™, AT&T HMIHY)
User: "Uh…we want to move…we want to change our phone line from this house to another house"

DARPA
CALO Project

Keyword Spotting
(e.g., AT&T)
System: *"Please say collect, calling card, person, third number, or operator"*

Thanks to VIvian Chen, Asli Celikyilmaz, & Dilek Hakkani-Tur: http://deepdialogue.miulab.tw

+ An Early Vision

- Apple's Knowledge Navigator
  - The Knowledge Navigator Video was made for Apple-CEO John Sculley's EDUCOM 1987 keynote in six weeks on a $60,000 budget
  - https://www.youtube.com/watch?v=umJsITGzXd0
- Is the Knowledge Navigator Siri?
  - Based on the dates mentioned in the Knowledge Navigator video, it takes place on **September 16, 2011.**
    - **The date on the professor's calendar is September 16, and he's looking for a 2006 paper written "about five years ago," setting the year as 2011.**
  - **In October 2011, at the iPhone keynote, Apple announced**
    - **https://www.youtube.com/watch?v=AU2uhG5es2E**
  - **Siri, a natural language-based voice assistant, would be built into iOS 5 and a core part of the new iPhone 4S.**
  - ***So, 24 years ago, Apple predicted a complex natural-language voice assistant built into a touchscreen Apple device, and was less than a month off.***

2011 Siri: A breakthrough for speech

- What is SIRI?
  - from Roberto Pieraccini, Head of ICSI, at Mobile Voice 2012
    - Practically infinite vocabulary
    - Contextual language understanding – ANSWERS ... NOT LINKS
    - Voice access to calendar and contacts, help make reservations, gives answer on lots of things, including the meaning of life
    - Integrated within iPhone, freely available to everyone (who buys an iPhone)

+ Where did Siri come from

- July 2003: CALO: Cognitive Agent that Learns and Organizes
  - DARPA Selects SRI, along with 20 leading research organizations, awarding $22M over five years
  - Goal: revolutionize how computers support military and other decision-makers. It is considered to be the largest artificial intelligence project in history

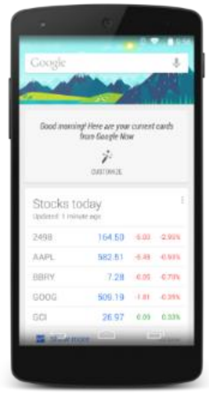
- December 2007: Siri is born
  - SRI spins off Siri, Inc. to bring the technology to consumers
  - October 2008
    - Siri, Inc. announces it has raised an $8.5 million
  - November 2009
    - Siri, Inc. raises a $15.5 million Series B financing round

- February 2010: Siri joins Apple
  - Siri, Inc. launches its Virtual personal Assistant app for the iPhone 3GS April 2010
  - Apple acquires Siri, Inc. from SRI

- July 2013: Siri comes to Boston
  - Apple opens a lab in Boston to work on speech recognition and Siri

+ From Siri to an explosion of devices

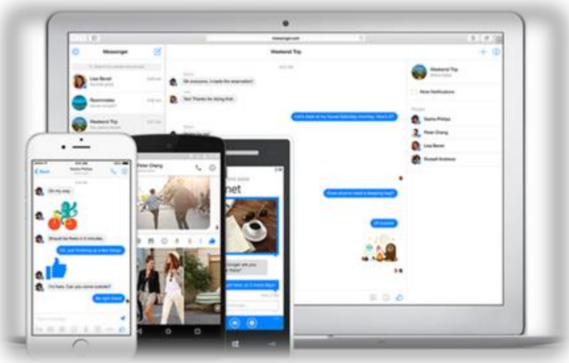
Apple Siri (2011)



Google Now (2012)
Google Assistant (2016)

Microsoft Cortana (2014)

Amazon Alexa/Echo (2014)



Facebook M & Bot (2015)

Google Home (2016)

Apple HomePod (2017)

Thanks to VIvian Chen, Asli Celikyilmaz, & Dilek Hakkani-Tur: http://deepdialogue.miulab.tw

+ What can these devices so for you?

- Get things done
  - set up alarm/reminder, take note
- Easy access to structured data, services and apps
  - find docs/photos/restaurants
- Assist your daily schedule and routine
  - commute alerts to/from work
- Be more productive in managing your work and personal life
  - Calendar, reminders,

+ Alexa “Skills”

- Business & Finance
- Communication
- Connected Car
- Education & Reference
- Food & Drink
- Games, Trivia & Accessories
- Health & Fitness
- Home Services
- Kids
- Lifestyle
- Local
- Movies & TV
- Music & Audio
- News
- Novelty & Humor
- Productivity
- Shopping
- Smart Home
- Social
- Sports
- Travel & Transportation
- Utilities
- Weather

+ Type of skills

One Shot

- Do it!
  - Set the timer
  - Turn the heat down
- Give me specific information
  - What time is
  - What’s the weather
  - What’s the capital of Oregon
- Tell me something from a category/library
  - Tell me a joke
- Walk me through some sequence
  - 7 minute workout
  - Play Jeopardy

Multiple turns

- Purchase something
  - 1-800-flowers
  - Order a pizza
  - Plan a trip
- Explore a space
  - Real estate
  - Movies
  - ??
- ??

+ Student applications

2014

- B-improved
- Jeeves
- FridgeBay
- RAMA: Rose Art Museum application

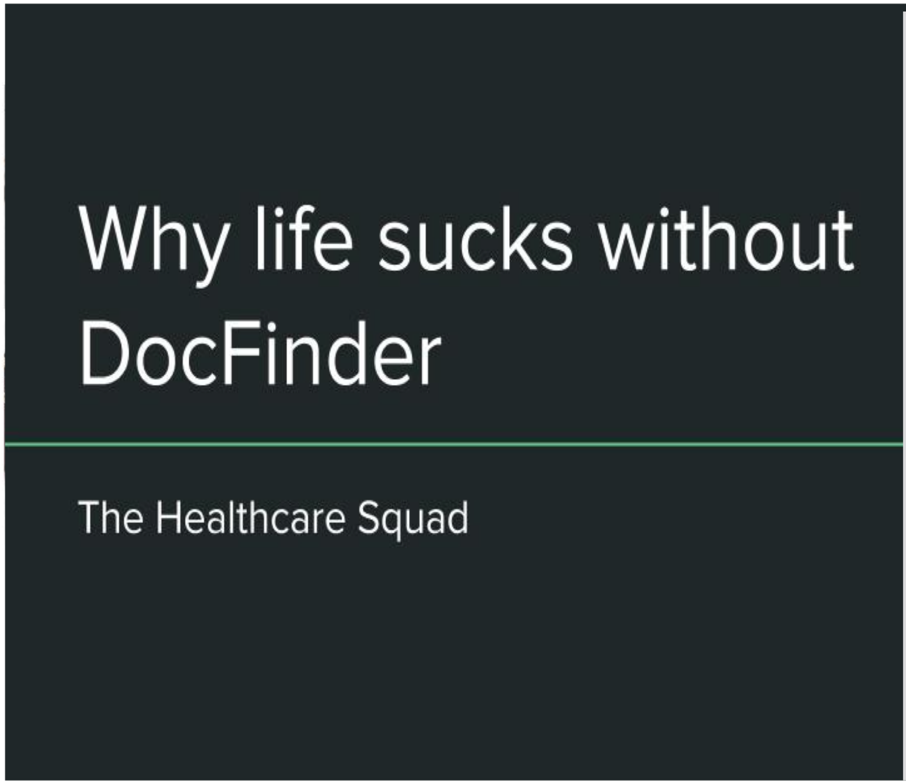
2015

- Bark!
- Memory Chess
- DiscoverDeis
- Workout

2016

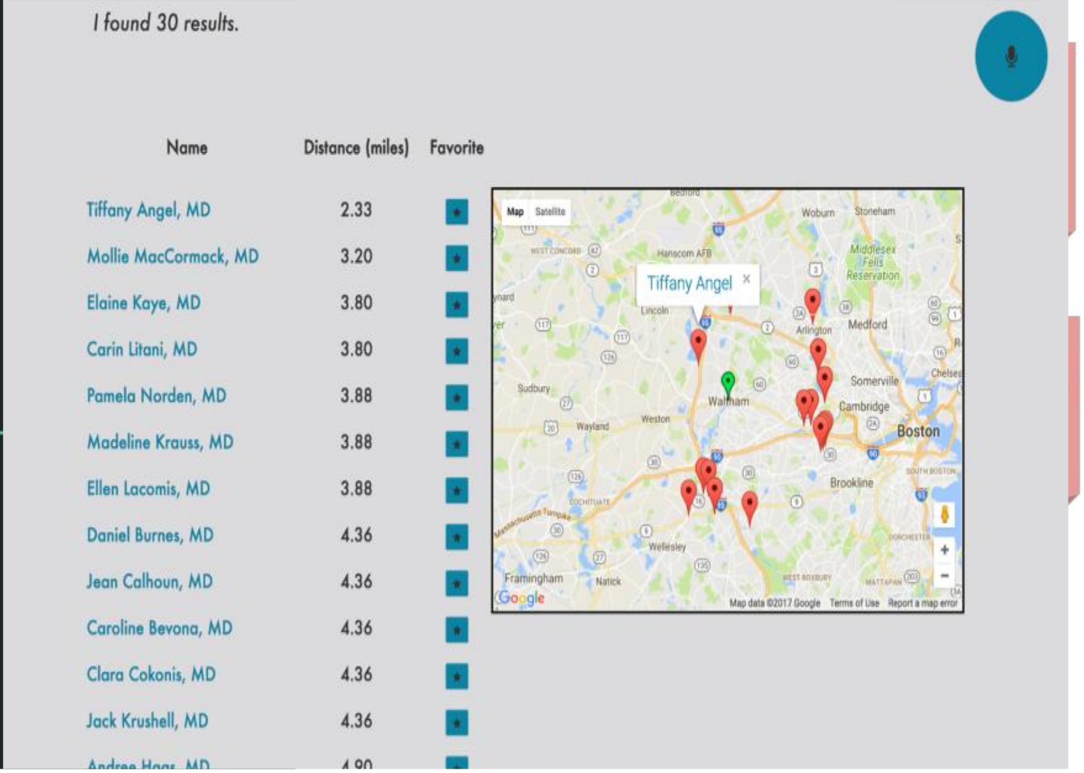
- Chef’s assistant
- Virtual Pet
- PlanDeis
- Language App
- Travel App
- Personal planning

JBS 2017



Why life sucks without DocFinder

The Healthcare Squad



I found 30 results.

Name	Distance (miles)	Favorite
Tiffany Angel, MD	2.33	★
Mollie MacCormack, MD	3.20	★
Elaine Kaye, MD	3.80	★
Carin Litani, MD	3.80	★
Pamela Norden, MD	3.88	★
Madeline Krauss, MD	3.88	★
Ellen Lacomis, MD	3.88	★
Daniel Burnes, MD	4.36	★
Jean Calhoun, MD	4.36	★
Caroline Bevona, MD	4.36	★
Clara Cokonis, MD	4.36	★
Jack Krushell, MD	4.36	★
Andrea Haas, MD	4.90	★

+ The AVIOS Student Speech Application

- Demonstrate your creativity and programming skills in voice-enabled and multimodal applications by entering the AVIOS Speech Application Development Contest organized by the Applied Voice Input Output Society.
- Develop a speech mediated application by <TBD, last year was in January> and win cash, prizes as well as world-wide recognition on the AVIOS web site and other public announcements.
- **DocFinder** from JBS2017 came in **First**!
  - All expense paid trip to Conversational Interactions Conference in San Jose
  - Cash prize!
- http://www.avios.org/

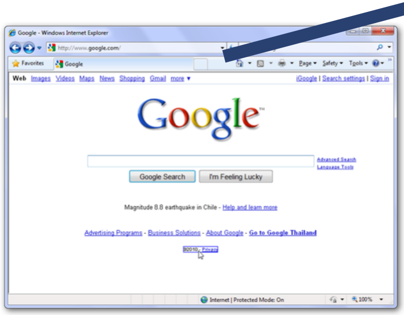
+ Why use language?

- More natural and convenient
- More efficient
  - Typing: 40 words/min
  - Speaking: 120 words/min
  - Reading: 80 words/min
  - Image (Is a picture really worth 1000 words?)
  - Brain: ?? Thoughts/min
- Devices are getting smaller
  - What does this do to typing speed? Reading ability?
- 3.65B Active unique mobile users

+ GUI vs. VUI

- Web-based application: What you see is your world
- Impact on programming: All the “context” is on the page

Type a URL



Type a search query, Select a link from results

Select a link from “history”

Select a link on a page

Select a link on a page

Welcome Marie!

You have 3 items in your shopping cart.

Log In provides profile and some context such as the shopping cart.

+ GUI VS. CUI (conversational UI)

	Website/APP's GUI	Msg's CUI
Situation	Navigation, no specific goal	Searching, with specific goal
Information Quantity	More	Less
Information Precision	Low	High
Display	Structured	Non-structured
Interface	Graphics	Language
Manipulation	Click	mainly use texts or speech as input
Learning	Need time to learn and adapt	No need to learn
Entrance	App download	Incorporated in any msg-based interface
Flexibility	Low, like machine manipulation	High, like converse with a human

+ Challenges

- Variability in Natural Language
- Robustness
- Recall/Precision Trade-off
- Meaning Representation
- Common Sense, World Knowledge
- Ability to Learn
- Transparency

+ From speech to things to …

- Jibo
- VUI
- Discourse.ai
- Google Cloud Platform
- IBM IoT
- SemanticMachines

+ High level design elements

- What platform?
  - Only smartphones?
  - Only ”smart speakers”
- Where’s the speech
  - Onboard? In the cloud?
- How does your proposed functionality align with the back end source of information?
  - Can’t “name” your bank accounts if the bank doesn’t track that info
- Are there other systems the app needs to integrate with?
- Is there a log in?
  - Can there be a user profile?
- Is any of the information being exchanged sensitive
  - personal, financial

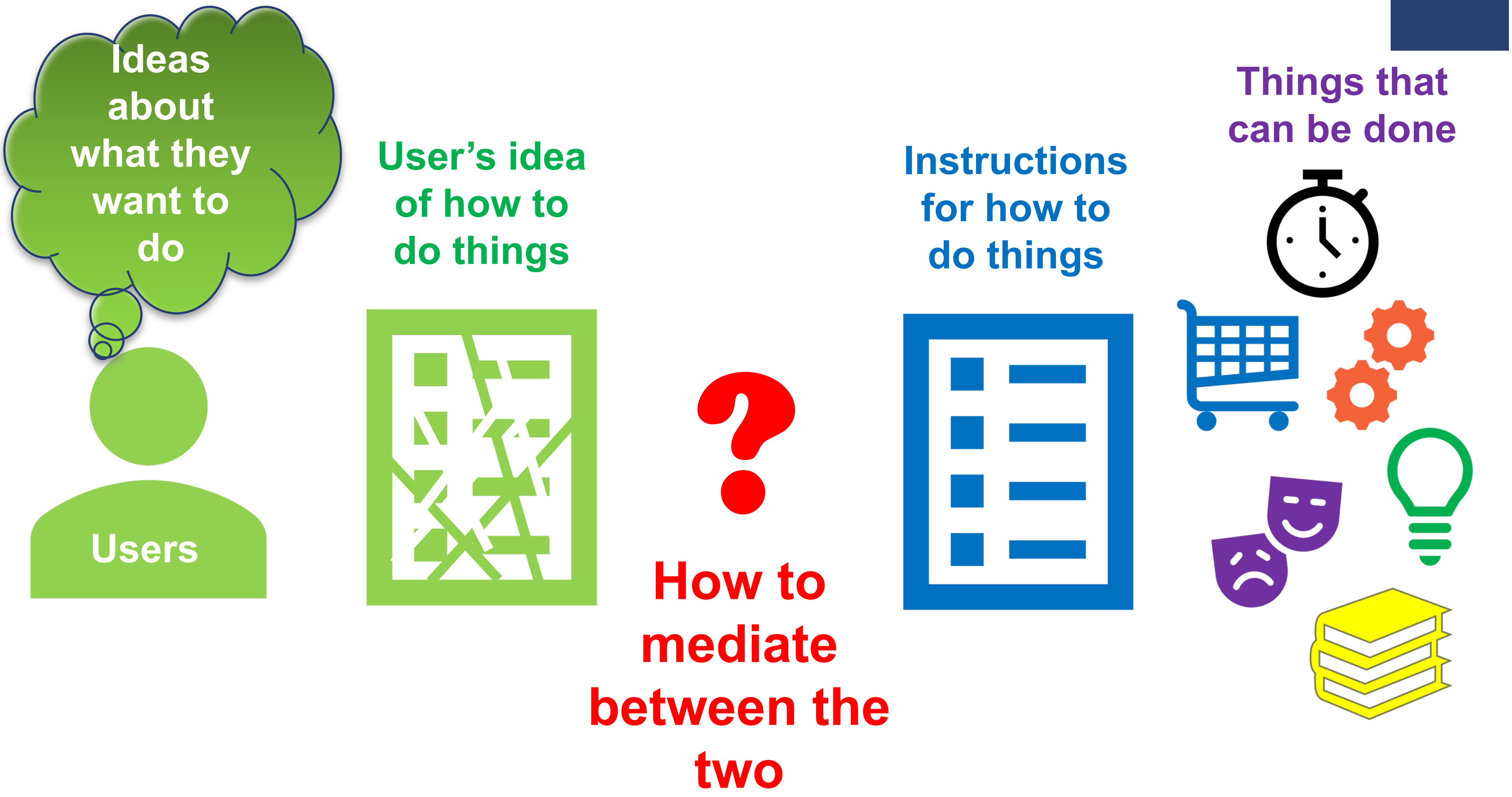
Dimensions of Voice Applications

- Push-to-talk vs. Voice trigger

- Communication modalities
  - Multimodal
    - Input and output can be audio/visual/tactile
  - Voice only
    - No visual interface—back to the “telephone” model
  - Text only
    - Chatbots
      - Can send a link, but interaction is only text

- Mixed initiative
  - User can change the topic or revisit previous dialog elements
  - System can take control to get more clarifying information
  - User can say anything at any point and get some intelligent response

+ Where to start?



Ideas about what they want to do

Users

User's idea of how to do things

How to mediate between the two

Instructions for how to do things

Things that can be done

+ Connecting the two

“Intent”:
Something a user wants to do

Set the timer

for 10 minutes

“Entity”:
Specifics required for that action

“Fulfillment”:
Code or query required to execute the intent

<Set + 10>

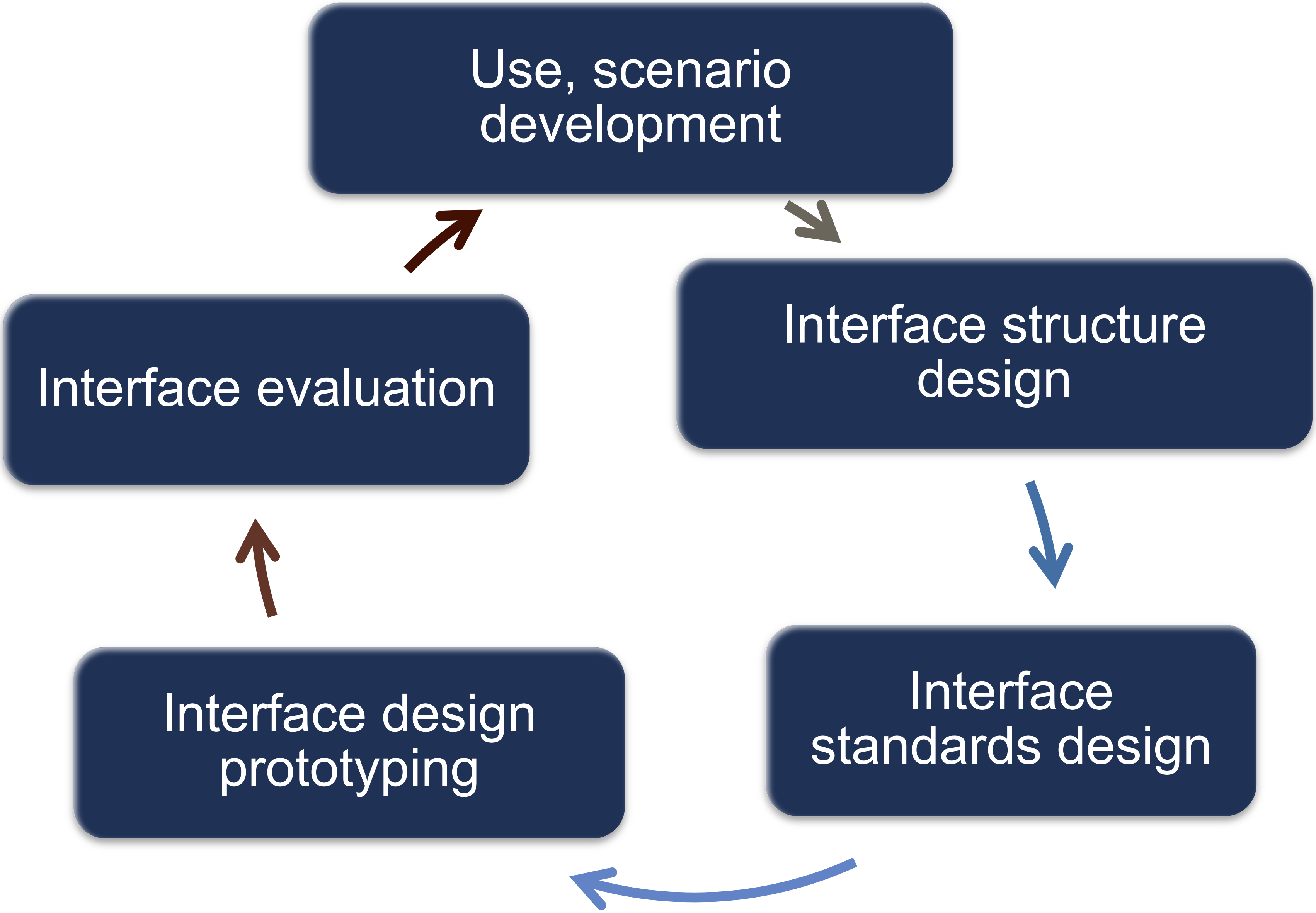
Timer set for 10 minutes starting now

“Response”:
What to say back to the user

+ First Blog Assignment

- Speech application review
  - Due SoC (Start of Class) Friday June 10.
  - Submit via Slack
- Select Two speech application and try it out
- Identify what worked and what didn't, how easy it was to use, how useful it was.
- Write a short review describing the application (functionality, platform) and how well it works (usefulness, limitations) and
- Post it to the class on Slack.
- Read the reviews from other classmates and comment on how the apps they reviewed compares to yours, whether you've used anything similar, or ask for more information about the app or its performance.

+ VUI Designer Bob Morse: Design cycle



Use, scenario development

Interface structure design

Interface standards design

Interface design prototyping

Interface evaluation