Course Overview

- Syllabus, schedule, assignments, papers available at cs.brandeis.edu/~cs115

- Morning classes: Dialog and Design

- Afternoon classes: Speech systems, NLP, and architecture

- Assignments:
  - Quizzes on vocabulary and readings
  - Occasional “Blog posts” on speech, apps, and other thoughts
    - Counts as class participation
  - Programming and presentation assignments
| Week 1        | Come up with ideas  
               | Get to know your classmates |
|---------------|---------------------|
| Week 2        | Share more ideas    
               | Broaden ideas and narrow scope  
               | Team building |
| Week 3        | Coalesce into 5 groups, each with a project idea  
               | 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
- Piazza: Forum for discussion
- Dropbox: Sharing data and code for class assignments
- Googledocs: Sharing documents
- Github: Storing and sharing programs
- Lots more specific to speech and NLP
  - Mashup speech server
  - CMU Pocketsphinx
  - NIST scoring softward
  - WIT.ai and API.ai NLP tools
+ Broad goals

- “Experiential 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
Exploring the space

- Learning how to think about the problem
- Learning what questions to ask to be able to answer the questions

**FINDING OUT ANSWERS BY ASKING THE RIGHT QUESTIONS**

- Form groups
- Go to Googledocs (check your email for the link)
- Select a scribe to type in the questions
+ **Question Formulation Technique (QFT)**

- *Four essential rules for producing your own questions:*
  - Ask as many questions as you can.
  - Do not stop to discuss, judge, or answer the questions.
  - Write down every question exactly as it is stated.
  - Change any statement into a question.
Our “Question focus”

Keyboards will disappear on mobile devices by 2020. Voice will be the input modality of choice.

IMPORTANT:
- We are not debating the statement
- We are brainstorming questions, which when answered, will help us have an informed discussion
“Questions” Exercise

- Now choose three questions that most interest your group
- Put a star next to the number in the first column of the spreadsheet for those questions
- Copy them to the “Top3” sheet of the Googledoc spreadsheet next to your group number
## Speech Applications

<table>
<thead>
<tr>
<th></th>
<th>Telephone</th>
<th>Microphone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive</td>
<td>• Call Routing</td>
<td>• Warehouse</td>
</tr>
<tr>
<td></td>
<td>• Call Center/IVR</td>
<td>• Command &amp; Control</td>
</tr>
<tr>
<td></td>
<td>• Directory Assistance</td>
<td>• Training (eg ATC)</td>
</tr>
<tr>
<td></td>
<td>• Voice search</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Smartphone apps</td>
<td></td>
</tr>
<tr>
<td>Non-interactive</td>
<td>• Voice mail transcription</td>
<td>• Dictation</td>
</tr>
<tr>
<td></td>
<td>• Call Center Analytics</td>
<td>• Court reporting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Multimedia search</td>
</tr>
</tbody>
</table>
What are “mobile voice applications”?

- **IVR**: “Interactive Voice Response:
  - Calling a company and getting a machine (and maybe a person later)

- **Mobile Applications**
  - Doing things you can do on a computer on a device you can carry around

- Are they really different?
What is SIRI?

from Roberto Pieraccini, Head of ICSI, at Mobile Voice 1012

- 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)
Siri: The Story So Far
(from Angel.com Mobile Voice presentation citing sirinotes.com)

- **July 2003**
  - DARPA Selects SRI to lead the Personalized Assistant that Learns (PAL) Program. SRI and a team of 20 leading research organizations are awarded $22M over five years to develop software, dubbed CALO to revolutionize how computers support military and other decision-makers. It is considered to be the largest artificial intelligence project in history.

- **December 2007**
  - SRI spins off Siri, Inc. to bring the technology to consumers.

- **October 2008**
  - Siri, Inc. announces it has raised an $8.5 million Series A financing round, led by Menlo Ventures and Morgenthaler Ventures.

- **November 2009**
  - Siri, Inc. raises a $15.5 million Series B financing round from the same investors as in their previous round plus Hong Kong billionaire Li Ka-shing.

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

- **July 2013**
  - Apple opens a lab in Boston to work on speech recognition and Siri.
More on Siri from Roberto

- Why is Siri successful?
  - Perception of intelligence
  - Fun to use it, witty, catchy personality
  - iPhone design and Apple marketing
  - Works relatively well for a certain number of tasks
  - Improves with time
  - So ... is speech recognition a solved problems?

- NO...and language understanding is even *less solved*.
  - Fails where humans don’t
  - Little basic science
  - More data, more improvements ... but the rate of improvement is diminishing
  - Looks like we are hitting the intrinsic limitations of the underlying models
  - Each new task requires almost the same new level of effort
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](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](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.
Application types

- Type of interaction/goal
  - Search for information
  - Accomplish a task (e.g. reservation)
  - Log information for later retrieval (e.g. calendar)
  - ??

- Origin of the app (or maybe who is it intrinsically tied to)
  - Personal assistant
    - Focus is you
    - Needs APIs to get anything done
  - Enterprise application
    - Focus is a specific enterprise, e.g. your bank, insurance company
    - Can control access and presentation of data
## Types of IVR Dialogs (Enterprise apps)

<table>
<thead>
<tr>
<th>Definition</th>
<th>Characteristics</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>One step from touch tone</td>
<td>Single words</td>
<td>S: Say or press one</td>
</tr>
<tr>
<td>Directed Dialog</td>
<td>Finite state grammars</td>
<td>S: What city are you leaving from? C: Boston</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S: What city are you flying to? C: New York City</td>
</tr>
<tr>
<td>Responses with multiple pieces of information</td>
<td>Multi-slot/Multi-token speech, Form filling</td>
<td>S: What city are you leaving from? C: I need to fly from Miami to Boston on February 6th</td>
</tr>
<tr>
<td>Open-ended prompts</td>
<td>Natural language call routing, Call steering</td>
<td>S: How may I help you? – C: A charge on my bill that – umm – I don’t understand</td>
</tr>
</tbody>
</table>
Dimensions of Voice Applications

- Push-to-talk vs. Voice trigger

- 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

- Multimodal
  - Input and output can be audio/visual/tactile
Characteristics of an “Advanced Dialog System”

- **Context aware**
  - “Remembers” what the user has already said and uses that information
  - Recognizes the users goals based on what has happened so far
  - Tracks the users “focus” so speakers can refer back to objects already mentioned

- **“Natural” interaction**
  - variability of expression
  - humor

- **“Self aware”**
  - Knows when to ask for clarification and when to move forward

- **Helpful**
  - Offers additional information or incorporates multiple steps
“Question focus” Revisited

- Keyboards will disappear on mobile devices by 2020. Voice will be the input modality of choice.

REMEMINDER:
- We are not debating the statement
- We are brainstorming questions, which when answered, will help us have an informed discussion
+ “Questions” Exercise

- Again choose three questions that most interest your group
  - These can be from the first set or the new ones

- Put a star next to the number in the first column of the spreadsheet for those questions

- Copy to “Top3 Round3” sheet on the Googledocs spreadsheet in the section labeled for your group
A look at the industry

- Mobile Voice 2014
  - http://mobilevoiceconference.com

- Focus is on commercial technology
  - Somewhere between a trade show and an academic conference
  - Lots of CTOs, research directors

- We can use these presentations as a window into the industry
  - http://avios.org/?page_id=5403
MV2014 Presentations

- Review Jordan Cohen’s presentation

The AVIOS Student Speech Application
January 2016

- 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.

- A group from JBS2014 came in 2nd last year!

- http://www.avios.org/
First Blog Assignment

- Speech application review
  - Due SoC (Start of Class) Thursday June 3.
  - Submit via Piazza

- Select a 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 Piazza.

- 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.