Building a voice enabled application
A study in 9 weeks

Marie Meteer
Brandeis University
Brandeis “JBS” intensive summer program

The context:
- Undergraduates with at least 1 year of programming
- 9 weeks, M-F 10am – 3pm+
- Since 2014
- 15 – 20 students

The question:
- What do you teach to get student groups from ideas to applications?
## Student applications

<table>
<thead>
<tr>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-improved</td>
<td>Bark!</td>
<td>Chef’s assistant</td>
</tr>
<tr>
<td>Jeeves</td>
<td>Memory Chess</td>
<td>Virtual Pet</td>
</tr>
<tr>
<td>FridgeBay</td>
<td>DiscoverDeis</td>
<td>PlanDeis</td>
</tr>
<tr>
<td>RAMA: Rose Art</td>
<td>Workout</td>
<td>Language App</td>
</tr>
<tr>
<td>Museum application</td>
<td></td>
<td>Travel App</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Personal planning</td>
</tr>
</tbody>
</table>
Voice, Web and Mobile Applications

- First 5 weeks
  - Web Application Development
  - Spoken Dialog and Design

- Final 4 weeks
  - Incubator → Showcase
Web application development

- Javascript
- Meteor
- MongoDB
- Agile Software Development
- HTML
- CSS
- Git
- Blaze

Agile Development

Next Iteration

Demo Release

Make Changes

Integrate and Test

Integrate and Test

Develop Functionality N

Develop Functionality 1

Develop Functionality 2

Develop Functionality 1

Integrate and Test

Integrate and Test

All Functionality Complete

Yes

No
Spoken Dialog Design

- Speech recognition
  - Understanding the challenges
  - Run recognition and score word error rate
- NLP
  - Entities and intents
  - Cloud services: api.ai and wit.ai
- Discourse
  - Grounding
  - Grician Maxims
  - Clarification
- Design and Usability
  - Personas
  - Scenarios
Basic Boxology

Voice & Data

Speech Recognition

Natural Language Understanding

Speech Synthesis (TTS)

Natural Language Generation

Dialog Manager

Context update

Response Strategy Selection

Task Manager

Voice & Images
How much and how to teach?
Design axioms:
Juhan Sonin at Involutions Studios
VUI Designer Bob Morse: Design cycle

- Use, scenario development
- Interface structure design
- Interface standards design
- Interface design prototyping
- Interface evaluation
High level design elements

- What platform?
  - Only smartphones?

- 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
Dialog Design

- Application description

- Personas
  - Short descriptions of different kinds of people who might use the application

- Use cases
  - Specific sequences of actions

- Tasks
  - What are the specific things a user can do
  - What information does the underlying application need to know to execute those commands
  - What are the “subtasks”, or steps required to complete the task
Users

- Who are the target users?
  - Create a “persona” for each user population

- Will there be naïve vs. expert users?

- Is this a “one time” app or will there by regular, repeated use?

- Is there a particular jargon or vocabulary involved
  - sports, financial services

- What is the level of attention of users
  - Will they use the app when driving?

- What are the environment the users will be in
  - Office, car, on the street
Rebecca is an 17-year-old high school senior looking at colleges in the Boston area. She’s interested in Brandeis, especially the biology programs, and wants to learn more. Rebecca is always on her iPhone but feels a bit awkward speaking to her phone in public but even more awkward speaking to admissions.

Prof. Ma will be a guest lecturer on campus. After his lecture, he has free time and would like to look around campus, visit the Rose and Slosberg, and meet other professors. He’s not the most technologically inclined person; he types slowly and is eccentric enough to not care about being seen talking to his iPad.
Scenarios: Workout app

- After a month of using the app Lee is starting to see progress and is feeling great. His adiposis is gone and feels that he has reached his goals. Lee stops working out for the next several days but the app challenges him to set new goals to exercise. Friends on social networking suggest some new goals for Lee.

- John is a 33 year old worker who is employed and works as a 9 to 5er in a busy office at a big company. He goes to the gym about 3 days a week and does the few exercises he knows. He spends most of his days at a desk working at a computer or in meetings where he still is in front of a computer. He uses his phone to always stay updated about work, and keep up with his friends and plan activities. He has recently decided that he wants to start seeing some progress and to step up his workout by learning new, more effective exercises and routines.
Scenarios → tasks

- Clear navigation instructions
- Customize Profile
- Some incentive/motivation to keep working out
- List of popular and effective exercises
- Ability to see and track progress
  - Setting goals and tracking them
- Strict exercise routine that motivates you to go to the gym or at the very least do a 15 min workout at home
- Find the app useful and helpful → Simple to use and navigate
  - Continue using it
  - Recommend to others
+ Intents and entities: Kitchen Helper

- Search under a category
  - “I want to make something Italian tonight”.
    - Intent: Find recipe
    - Entity: Italian, tonight

- Show results
  - “Show me the results”
    - Intent: show result
    - Entity: result
  - “Show me the second one”
    - Intent: show recipe
    - Entity: ingredients, instructions

- Search from favorites
  - “I want to make the mac and cheese recipe in my favorites”
    - Intent: Find recipe from favorites
    - Entity: Recipe in favorites

- Show results
  - “Show me the recipe”
    - Intent: look at recipe
    - Entity: ingredients, instructions
Usability tests: Personal Planning

Scenario
- You are a freshman engineering student at CalTech and desperately want to stay organized. You create a school category and enter in the time and locations of all of your classes. For your morning English class, you want to create a “text” to take notes, since this is the only class you are taking this semester in which hand-written notes are preferable.

Feedback Questions:
- What would you like to see that you didn’t see?*
- What would you like to change about the experience?
- How often would you like to use this app?
- What features of the app are you most likely to use?*
- Is there any feature in the app that confuses you?*
- What are some features that you are not very comfortable with?
- Would you be willing to pay for this app? If so, what’s your pricing limit?*
- Which feature of this app did you think outperformed other similar apps?*
Speech Recognition

- A dip into the process: The “Pizza” assignment
  - Write a grammar to order a pizza
  - Record 10 pizza orders
  - Upload to Interactions “Mashup”
  - Run recognition
  - Use NIST ScLite to get a word error rate
  - Combine audio from the entire class
  - Repeat recognition
  - Improve grammar, repeat
NLP in the cloud: AIP.ai, wit.ai

- “Understanding” based on
  - **Intents**: What is the speaker trying to do with this utterance
  - **Entities**: What specific entities are being included in the utterance
- Provide a way for the NLU component to learn from examples
  - Define a set of possible intents
  - Define a set of entities
  - Create a corpus of sentences and annotate each with the intent and entity

- Advantages
  - Really **easy to use**: No need to understand formalisms
  - Can take advantage of other applications to learn similar goals
- Disadvantages
  - Focus is on individual utterances
  - No way to handle discourse segments where intent and entities might be spread across multiple utterances (progress being made)
Challenges

- Speech
  - Unusual constructions: Pawn to king’s knight four
  - Brandeis building names: “Gerstenzang”
  - Course abbreviations: “Psych” is ok, but “Anth” doesn’t work

- Context
  - There are 3 buildings named “Shapiro”, but if I just asked about lunch, then I want the student center not the athletic facility

- Shaking the tyranny of the pages and menus
  - Go to the exercise page. What is my next exercise for today?
  - Go to my fridge list. What things do I have in my fridge?
Shameless Ad for Brandeis

- Brandeis Computational Linguistics MA program

- New: [blogs.brandeis.edu/compling](blogs.brandeis.edu/compling)

- “Meet and Greet” Industry Reception at Brandeis
  - Tuesday February 28
  - Meet our students, get to know our program
2016 Applications

- Chef’s assistant
  - https://drive.google.com/drive/u/1/folders/0B6z1otdq2OZuX1YyRUJFMEtpQ0E

- Virtual Pet
  - https://drive.google.com/drive/u/1/folders/0BxEvzlcrn6mWNTF6QmVyN3dCUG8

- PlanDeis
  - https://drive.google.com/drive/u/1/folders/0B-zCjBGx5719V3FsV2JKTE04TWM

- Language App
  - https://drive.google.com/drive/u/1/folders/0B92WGLWvR3vmMDFvTHIPRUYwWEE

- Travel App
  - https://drive.google.com/drive/u/1/folders/0ByeRtDqHrvKldFZjMkYxYkw4U1U

- Personal planning
  - https://drive.google.com/drive/u/1/folders/0B1bWj8AgU-NLX09weE1iZUdxQUE
Design

- Personas
- Scenarios
- Tasks
- Entities and Intents
- Usability testing
- Interface design