Speech Processing 11-492/18-492

Spoken Dialog Systems
Simple State Based Dialog Systems
State-based Dialogs

- **What are the objects in the domain?**
  - Write them down

- **Write down three easy example dialogs**
  - Don’t just think about it, write it down

- **Label each turn with what state you think they are**
  - Relabel them with your set of states
  - Change the dialog to match what you can support
  - Identify what you are not going to do
Simple state-based dialog systems

- Get Name
- Get Account number
- Get PIN
- Present balance
- Go back to start or exit
State-based Dialogs

- Get Name:
  - What is your name?
    - ASR Name
    - May be correct (in the database)
    - May be unknown (not in database)
    - May not be name (What do I say?/Help/Repeat)
    - Should you echo the recognized name?
      - Confirmation (or not)
State-based dialog

- **Get name**
  - Check in database
  - Ask again if not
  - Deal with help
- **Get account number**
  - Check in database (with name)
- **Confirm account number and name**
  - For security
State-based Interaction

- **Trees can get very large**
  - User can get lost easily

- **You want to minimize the number of turns**
  - Faster throughput means more calls
  - Faster throughput means happier customer
The level of help

- **First time users *need* a successful call**
  - Otherwise, they wont call back
  - *Having very helpful prompts is good*
    - At start, gets annoying quickly

- **Designing prompts is a craft**
  - What is spoken should be understood
  - How much should you tailor it to the user
How do users speak

- **What is your name?**
  - *My Name is …*
  - *My Name, its …*
  - ...
  - *I’m …*
  - *I’m called …*
  - *And a lot more*

- **Guess what a user might saw**
  - *Write it down*
  - *Do this 5 times.*
  - *Then do it another 5 times (really do that!)*
Speech Processing 11-492/18-492

Spoken Dialog Systems
Case-study: Personal Digital Assistants
Speech-based Personal Digital Assistant

- Build a speech enabled PDA
  - Speech in/out for individual use
- Goals
  - Control schedule
  - Control messaging
  - Replace personal assistant
- Any similarity to any existing product is purely coincidental

Disclaimer: Much of this is relevant to Apple’s Siri, but this information is general and may or may not be what is in Siri.
SPDA: Scope

- Schedule
- Calls (in and out?)
- Navigation
- Finding local businesses
  - With reviews
- Open questions
- Reminders/Alarms
“Call John”

“Call John, Bill and Mary and setup a meeting sometime next week about Plan B that’s fits my schedule”

“Make a reservation at a local Chinese restaurant for 4 at 8pm.”

“You should call your mom as its her birthday”

“I have sent flowers to your mom as its her birthday”
Cognitive Assistant that Learns Online
- DARPA project (2003-2008)
- Led by SRI (involved many sites, including CMU)

Personal Assistant that Learns (Pal)
- Answers questions
- Learn from experience
- Take initiative

Spin-off company -> SIRI
- Acquired by Apple in April 2010
SPDA: Platform

- Desktop
  - Computational power

- Phone (non-smartphone)
  - General Magic
    - Was handheld, became phone based
  - Led into GM’s OnStar

- Smartphone
  - Local to device
  - With Cloud
Smartphone + Cloud

- **Smartphone**
  - **Know about user**
    - Contacts, Schedule etc
    - Same speaker
  - **Some computation possible on device**

- **Cloud**
  - **Learn from multiple examples**
  - **Retrain acoustic/language/understanding models**
Voice Search and User Feedback

- **Voice Search**
  - Google, Bing, Vlingo, Apple

- **Get users to help label the data**
  - Listen to user
  - Show best options
    - They select which one is correct

- **Find out how users actually speak**
  - Full sentences vs “search terms”
  - How do English speakers say ethnic names
Voice Search: Simplifications

- Too many words …
- Context
  - Where you are (location: home/not home)
  - What is on your phone (contacts)
  - What you’ve said before
Personality

- **Have a character**
  - Calls you by name (you choose)
  - *Pushy, helpful, nagging ...*
  - Allow user choice

- **Personalize it**
  - May form better relationship with it

- **e.g. Siri**
  - US and UK are female/male
Make it do things well

- **Targeted apps**
  - Chose what it will do well

- **Say, 12 different apps**
  - Have target (hand written) interaction
  - Chose what fields you need, and how to interact with the back end data
  - If all else fails dump result in Google

- **Hardware aid**
  - Infra-red detector for VAD
Make sure people know it's there

- (Voice search has been on PDA’s for years)
- Get a *lot* of people to use it
- Give “silly” examples
  - People will repeat them, you can adapt your system and expect them to say them
Know Your Users

- Young educated
- Standard English speakers
  - (Non-native too?)
- Can you train them to use it better
  - Get them to adapt
What is Missing?

- Add an SDK
  - Other app developers will want to allow speech
  - May make it harder to distinguish
- Dialog context
  - What was said in the previous utterance
- Others …
Will it work?

- Will people talk in public
  - Talking on the phone is now acceptable
  - Talking to the phone …

- Will people continue to use it
  - Cool at first, but easier to use menus
  - Only use for setting alarms

- Long term use …

- But others may join in anyway