Speech Processing 11-492/18-492

Spoken Dialog Systems
Tree based dialogs
VoiceXML
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 won't 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
VoiceXML

- A W3C standard for voice browsing
  - XML based “programming” language for speech
  - Output synthesized (and recorded) speech
  - Recognition of speech
  - Recording of spoken input
  - Telephony features
ASR

- From Grammars (JSGF: java speech grammar format)
- From tri-grams
- From “Domain Managers”
  - Credit card numbers
  - City, Stats
**TTS**

- `<ssml>` markup
- Choice of voice
- Choice of language
- Choice of how to pronounce things
- Specify breaks, timing, emphasis
<vxml version="1.0">
<meta name="author" content="John Doe"/>
<var name="hi" expr=""Hello World!""/>
<form>
 <block>
  <value expr="hi"/>
  <goto next="#say_goodbye"/>
 </block>
</form>
<form id="say_goodbye">
 <block>
  Goodbye!
 </block>
</form>
</vxml>
Basic Tags

- `<form id="xxxx">`
  - `<goto next="#xxx">`
- `<field> gather info from user through speech`
- `<record> record data user`
- `<subdialog> performs some sub dialog`
<form id="getBusNumber">
  <field name="BusNumber">
    <prompt>Which bus line do you want?</prompt>
    <grammar src="grams/bus.gram">
      <help>Please say your desired bus number, e.g. 61C</help>
    </grammar>
  </field>
</form>
Flow of Control

- **Goto**
  
  `<goto next="#GetBusNumber">`
  `<goto next="Trains.vxml">`

- **<if cond="BusNumber == '501'">**
  
  `<prompt> Sorry that bus no longer runs</prompt>`

- **<elseif cond="BusNumber == '56U'">**
  
  `<prompt> Sorry it’ll be a long wait</prompt>`

- **<else />**
  
  `<prompt> One will be along shortly</prompt>`

`</if>`
<var name="var1" expr="hello"> 
<prompt> I just wanted to say <value expr="var1"></prompt>
</value> 
<assign name="var1" expr="goodbye">
Speech Recognition Grammar Specification

- (SRGS)

Augmented BNF

$order = I \text{ would like a } drink$

$drink = coke \mid pepsi \mid mountain\_dew$
VoiceXML Browsers

- **Compatibility**
  - Not as compatible as one would like
  - `<objects>` can be different (but useful)
    - City, State recognizers
  - ECMAScript (Javascript)
Beyond VoiceXML (in VoiceXML)

Mixing html/cgi scripts in VoiceXML

- Use php to generate VoiceXML files
- Use urls (with ?...) to calculate/get data
  - http://weather.com?zip="15213"
- Use urls to get waveforms
  - http://tts.com?text="Hello World"
VoiceXML future

- **N-gram grammar Markup Language**
  - Many browsers have own extensions

- **Pronunciation Lexicon Markup Language**
  - A way to add new items to the lexicon
  - Hard to find good standards

- **Call Control Markup Language**
  - For management and logging of calls
Microsoft SALT

- **SALT tags**
  
  Listen DTMF prompt bind grammar (plus ssml)

- **Designed for desktop not just phone**

- **Design to be shared documents**
  
  - Viewing (HTML) and Speech (SALT)
Available Systems

- Nuance
- Be-vocal
- Tell Me
  - Tell-me studio
- OpenVXI/publicvoicexml.org
- HALEF
- Many others