• A Recap
• Where Are Dialog Systems Now?
• Application Complexity
• Call Flow Notation
A Recap: What You’ve Covered So Far

• The Telebet System: A Case Study of A Successful Voice App
• What’s Involved in Dialog Design and Prompt Design
• The W3C Speech Interface Framework
  – VoiceXML 2.0, 2.1, and VoiceXML Development Platforms
  – Using JavaScript with VoiceXML
  – Grammars, Speech Synthesis and Semantics in VoiceXML
  – Mixed Initiative in VoiceXML
• A Real Estate Application
• Other Topics: SALT, Speaker Verification, Voice Over IP
Typical Exam-like Questions

• What are the properties of a successful voice application? An unsuccessful voice application?
• Examples of dialog design, grammar design and prompt design
• VoiceXML specifics
• What are the pros and cons of VoiceXML vs SALT?
• How does speaker verification work? What are its limitations?
• How does Voice over IP work?
What’s Ahead

- Assumptions:
  - I assume you fully understand VoiceXML
  - I assume you have fully understood Rolf’s points re dialog and prompt design
  - I expect high level participation in class

- Coverage:
  - Understanding the limitations of today’s dialog systems
  - Understanding how we can advance today’s technology
Outline

• A Recap
• Where Are Dialog Systems Now?
• Application Complexity
• Call Flow Notation
Where Are We Now?

• “IVR [interactive voice response] technology is at a point now where consumers almost cannot tell the difference between talking to a person and talking to a computer.” [Richard Feinberg, director of the Purdue University Center for Customer Driven Quality, quoted in the online *CRM Daily*, February 18th, 2004]
The Architecture of an SLDS
What an SLDS contains

- **Speech Recognition**—analyses the audio speech input signal to extract linguistic units such as words or phonemes
- **Language Understanding**—determines the meaning of the input
- **Dialog Management**—manages the flow of the conversation, maintaining history and context, directing its course, accessing the database, and formulating responses
- **Database**—stores the information which provides the dialog content
- **Language Generation**—puts the responses into words
- **Speech Synthesis**—produces the audio speech output signal
Typical Systems Today

- Speech Recognition: accuracy of 85-90%+
- Speech Synthesis: high quality for short utterances
- Dialog Management: strongly-managed dialog flow
- Language Processing: Very limited
- Language Generation: Very trivial
Where We Are Headed?

• "The aim is to put back some of the intelligence which existed in the network 50 years ago when a user simply lifted the handset and asked to be connected to the service or number required." [Wyard et al (1996), ‘Spoken Language Systems: Beyond Prompt and Response’, *BT Technology Journal*.]
Pushing the Envelope

- Basic Speech Recognition accuracy unlikely to advance much further with current technologies
- Dialog Management techniques need to be more flexible
  - Mixed-initiative
  - Open vocabulary
- Language Processing can be much more sophisticated
- Language Generation can be much more sophisticated
Current Systems Case Studies

- Telstra Directory Assistance
- Taxis Combined Taxi Booking
A Sample Dialog

[Caller dials 1223]

TDA: Telstra Directory Assistance. What name, please?

User: Qantas.

TDA: Answering yes or no, did you request Qantas?

User: Yes.

TDA: The number is 13 13 13.
A More Complex Dialog

[Caller dials 1223]

TDA: Telstra Directory Assistance. What name, please?
User: Centrelink.

TDA: Answering yes or no, did you request Centrelink?
User: Yes.

TDA: Please choose Austudy or … or say the department you require.
User: Austudy.

TDA: …
What's Involved: The Business Case

- Total number of calls to Directory Assistance is around 8 million per week
- Automating just 10% of calls with around 90% accuracy justifies cost of installation of the technology
- A small proportion of requested numbers account for a high proportion of all requests
- Around 140 numbers account for 10% of calls
Cumulative Counts for Calls

Call Volumes for the Most Requested Directory Assistance (DA) Names

Call Volumes as a Percentage of the Total Number of DA Calls per Week

where \( n \) is the \( n \)th Most Requested DA Name
The Importance of Careful Data Analysis

• Implication is that 140 requests will account for 10% of calls
• But:
  – 140 is the number of provided numbers
  – This is not the same as requested names
• For example:
  – Kingsford Smith Airport
  – Sydney International Airport
  – The airport in Sydney
  – …
## The Devil is in the Details

<table>
<thead>
<tr>
<th>Dialled Number</th>
<th>Frequency (calls per week)</th>
<th>Accumulative Frequency (calls per week)</th>
<th>Accumulative Frequency as a percentage of total calls per week</th>
</tr>
</thead>
<tbody>
<tr>
<td>131319</td>
<td>78,934</td>
<td>78934</td>
<td>1.0%</td>
</tr>
<tr>
<td>132203</td>
<td>37,502</td>
<td>116436</td>
<td>1.5%</td>
</tr>
<tr>
<td>1225</td>
<td>30,450</td>
<td>146886</td>
<td>1.8%</td>
</tr>
<tr>
<td>132200</td>
<td>29,005</td>
<td>175891</td>
<td>2.2%</td>
</tr>
<tr>
<td>12550</td>
<td>24,991</td>
<td>200882</td>
<td>2.5%</td>
</tr>
<tr>
<td>12552</td>
<td>20,437</td>
<td>221319</td>
<td>2.8%</td>
</tr>
<tr>
<td>132861</td>
<td>17,550</td>
<td>238869</td>
<td>3.0%</td>
</tr>
<tr>
<td>132265</td>
<td>17,539</td>
<td>256408</td>
<td>3.2%</td>
</tr>
<tr>
<td>12454</td>
<td>17,255</td>
<td>273663</td>
<td>3.4%</td>
</tr>
<tr>
<td>1300300937</td>
<td>15,964</td>
<td>289627</td>
<td>3.6%</td>
</tr>
<tr>
<td>132850</td>
<td>14,396</td>
<td>304023</td>
<td>3.8%</td>
</tr>
<tr>
<td>131008</td>
<td>14,252</td>
<td>318275</td>
<td>4.0%</td>
</tr>
<tr>
<td>132213</td>
<td>13,877</td>
<td>332152</td>
<td>4.2%</td>
</tr>
<tr>
<td>131314</td>
<td>13,753</td>
<td>345905</td>
<td>4.3%</td>
</tr>
<tr>
<td>131313</td>
<td>12,457</td>
<td>358362</td>
<td>4.5%</td>
</tr>
<tr>
<td>180181111</td>
<td>12,332</td>
<td>370694</td>
<td>4.6%</td>
</tr>
<tr>
<td>283328888</td>
<td>11,310</td>
<td>382004</td>
<td>4.8%</td>
</tr>
</tbody>
</table>
Final System

- About 1600 top level names, covering 15% of calls
- Multiple ways of saying each name
- Some names require second questions
- Some delivered numbers are provided in response to a very wide range of customer enquiries → not automatable
The Name Grammar: Top Level

NAME
[
2WS_SUB~0.001312147
3AW_SUB~0.002891474
3M_AUSTRALIA_SUB~0.002072124
6IX_SUB~0.001169652
6PR_SUB~0.001169652
92.9PMFM_SUB~0.001656512
94.5FM_SUB~0.001852443
A_B_C_SPARE_PARTS_SUB~0.001211213
A_N_A_HOTEL_SUB~0.005397023
AAMI_SUB~0.035600162
AAPT_SUB~0.006471677
ABC_RADIO_SUB~0.01284242
ABC_TAXIS_SUB~0.003117092
ABC_TELEVISION_SUB~0.004607359
... ]
The Name Grammar: Name Expansions

3AW

[ (three a_w) (three a_w radio) (radio three a_w) ]
{<name "3AW">}

AAMI

[ ([a_a_m_i (double a_m_i)] ?insurance) (double letter_a m [i y] ?insurance)~0.001 (letter_a letter_a m i ?insurance)~0.001 amy (amy insurance) (amy insurers) (australian associated motor insurers) (a_a_m_i insurance) (a_a_m_i insurers) ]
{<name "AAMI">}

Comp349 Week 07
Was it a Success?

- From the business case perspective, yes
- From the public relations perspective, perhaps not
Consumer Misunderstandings

An Example:

• "Jon Fain on ABC local is grilling the Telstra manager for the new directory assistance 'service' which is voice recognition, not punch letters coding. It failed abysmally when the local news reader asked for the Metropolitan Ambulance Service and Peter McCallum Hospital, with the example given was played right after the guy said it was primarily for accessing common places, which in Melbourne would definitely include these two places. ..."
Current Systems Case Studies

- Telstra Directory Assistance
- Taxis Combined Taxi Booking
A Sample Dialog

User calls 8332 8888
System: To book a taxi, say yes. To check a booking, say check. For anything else, say no.
User: Yes.
System: Are you travelling from 14A Spears Avenue Balmain?
User: Yes.
System: And what suburb are you going to?
User: Milson’s Point.
System: How many passengers are travelling?
User: One.
System: Are you ready now?
User: Yes.
System: Confirming your booking for a taxi from 14A Spears Avenue Balmain to Milson’s Point for one person leaving now. Is that correct?
User: Yes.
System: Your booking for the next available taxi is confirmed. Your reference number is . . .
Why This Works

• Only handles calls from landlines, so can use calling number to determine calling location
• Only asks for destination suburb, not the complete address
• Asks questions that assume sensible defaults: 'Are you ready now?'
• Takes less than 1 minute to book a taxi, with no waiting in a queue for a human agent
• Automates somewhere upwards of 70% of calls, and transfers to an agent at first indication of problems
A Grammar for 'Airport'

```plaintext
( airport { <SuburbCode "AI2"> return("airport") })  ~1.13
( the airport { <SuburbCode "AI2"> return("airport") })  ~1.13
( the domestic terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( the international terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( the domestic airport { <SuburbCode "AI2"> return("airport") })  ~1.13
( the international airport { <SuburbCode "AI2"> return("airport") })  ~1.13
( qantas domestic terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( the qantas domestic terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( qantas international terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( the qantas international terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( ansett domestic terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( the ansett domestic terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( ansett international terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( the ansett international terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( impulse domestic terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( virgin domestic terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
( virgin blue domestic terminal { <SuburbCode "AI2"> return("airport") })  ~1.13
```
What People Really Say

System: And what suburb are you going to?
Caller: oxford street
taylor square ohh no darlinghurst
chinatown
i don't know
what suburb are we going to
airport where the hell's the airport
balmain and then to the city
first stop to surry hills and then second stop to kingsgrove
look tomorrow morning i want to book for
i'm going to middle street to doctor smith's the eye surgeon
but i'm not going to a suburb
Is it a Success?

• For the caller, avoids waiting in queues; takes 1 minute to book a taxi
• For the business, automates 70%+ of calls
Some Conclusions

• You need a strong business case to justify speech recognition, but a business case is not enough: you also need to persuade the customer that the solution is a good one

• You don’t need to automate all the calls, but it’s a good idea to let the caller know which ones you do automate

• Elegant solutions are often simple and exploit non-ASR aspects of the context
Outline

- A Recap
- Where Are Dialog Systems Now?
- Application Complexity
- Call Flow Notation
Determining Application Feasibility

- Before committing to designing and delivering a voice application, you need to assess the feasibility of building what the client wants.
- Application feasibility is determined by a number of factors.
- A key factor is application complexity.
Dimensions of Application Complexity #1: Vocabulary

• Does the application requires a small vocabulary or a large vocabulary?

• If it requires a large vocabulary, what problems might you face and how might you overcome them?
Dimensions of Application Complexity #2: Dialog Complexity

- How complex does the dialog need to be?
- What are the trade-offs between overall functionality and dialog complexity?
- What kinds of desirable functionality might you have to omit in order to make the application feasible and why?
Dimensions of Application Complexity #3: Dialog Initiative

- Does the application require more than a form-filling approach?
- What issues might arise in allowing a more mixed initiative dialog style?
Dimensions of Application Complexity #4: Syntactic Complexity and Grammar Coverage

- How broad would your grammars have to be to cover the likely user inputs?
- How will you determine the required grammatical complexity?
Dimensions of Application Complexity #5: Personalisation

- Does the application offer scope for system behaviour to be tailored to the individual user?
- How might you achieve this?
Dialog for Review #1: The Nuance Travel Planner

[S1] System: Welcome to Travel Plan, a service of Nuance Communications. What city would you like to leave from?


[S2] System: What city are you flying to?


[S3] System: Please say the destination city; for example, Atlanta Georgia.


[S4] System: On what date are you travelling?

[U4] User: Uh, on May the 28th. …
The Nuance Travel Planner, Continued

[S5] System: Around what time would you like to leave?
[U5] User: I've got to be in San Diego before 10am.
[S6] System: Please say your approximate departure time. For example, 10am.
[U6] User: Oh, probably 8am or so?
[S7] System: Please say your approximate departure time. For example, 10am.
[U7] User: Uh, 8am.
[S8] System: You are travelling from San Francisco to San Diego on Tuesday May 28th at 8am. Is this correct?
[U8] User: Yes, that's fine.
Dialog for Review #2:

**GUS**

[S1] System: Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go?


[S2] System: What time do you want to leave?

[U2] User: I must be in San Diego before 10am.

[S3] System: Would you like PSA flight 102 that arrives at 9:15am?

[U3] User: What is the next flight?

[S4] System: Air California flight 310 that leaves at 8:30am and arrives at 10:00am.

[U4] User: I'll take the first one.

[S5] System: In what name should I make a reservation, please?

[S6] System: I have confirmed the following flight: PSA flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 730am and arrives at 915am. What date do you want to return on?


[S7] System: Would you like the flight that leaves at 745pm?


[S8] System: I have confirmed the following flight: PSA flight 307 on Friday May 30 from San Diego to San Jose that leaves at 745pm and arrives at 930pm. Thank you for calling. Goodbye.
Dialog for Review #3:
The CMU Communicator

[S1] System: Hello. Welcome to the Carnegie Mellon Communicator. Are you a registered user?

[U1] User: Hi, this is Alex Rudnicky. I'd like to fly to Charleston, please?

[S2] System: Hello, Dr Rudnicky, I'm glad to hear from you again. You can ask me for help at any time. You can fly to Charleston in South Carolina or Charleston in West Virginia.


[S3] System: A flight to Charleston departing Pittsburgh, on what day?

[U3] User: On the 15th?

[S4] System: I'm sorry, I must have misheard the date. A flight on Tuesday, February 15th. Ok, what day would you be traveling?
[S5] System: Traveling on Wednesday March 15th. One moment please... let's see... flights from Pittsburgh to Charleston. ... I have a nonstop on US Airways departing Pittsburgh at 1135am arrives Charleston at 135pm.

[U5] User: How about something later?
[S6] System: The flight after that departs Pittsburgh at 210pm, arriving in Charleston at 410pm. Is that okay?

[S7] System: Will you return to Pittsburgh from Charleston?
[U7] User: Yes, on the Saturday?
Traveling on Saturday March 18th? Okay. I have a non-stop on US Airways departing Charleston at 1105am, arrive Pittsburgh at 113pm. Do you want to take that?

That's perfect, I'll take it.

Do you need any hotel or car reservations?

Yes, I'll need a car.

No hotels. Do you need a car in Charleston?

Yes please.

I've made a request for a car with Avis in Charleston. Do you want a summary of your trip?

Yes please.
The CMU Communicator, Continued

[S12] System: Your first flight is a US Airways flight 4072 leaving Pittsburgh at 210pm on Wednesday March 15th arriving in Charleston at 410pm. Then, the next flight is a US Airways flight 4120 leaving Charleston at 1105am on Saturday March 18th, arriving in Pittsburgh at 113pm. I've made no hotel reservations for your trip. I've requested a rental car in Charleston. Are you satisfied with this itinerary?


[S13] System: Mail sent to to alex.cmu.cs.edu. Thank you for using the Carnegie Mellon Communicator. Please tell me any comments. When you're finished, Just hang up.
Where Do These Applications Stand on Complexity?

- Vocabulary
- Dialog Complexity
- Dialog Initiative
- Syntactic Complexity and Grammar Coverage
- Personalisation
Outline

• A Recap
• Where Are Dialog Systems Now?
• Application Complexity
• Call Flow Notation
Call Flow Requirements

• Over the next few weeks you will be required to present a number of call flow designs in the tutorials
• You must submit these electronically ahead of time to rdale@ics.mq.edu.au
• You must use Microsoft PowerPoint
• You must follow the specified formatting conventions
Using PowerPoint for Call Flows

- Stick to the agreed conventions for the various different elements
- Label exit arcs from states with indicative user inputs
- Use AutoShapes for Basic Shapes and Connectors to allow easy rearrangement
- Manage complexity by encapsulating sub-dialogs
Call Flow Notation

• Speech recognition states:
  – Use rectangular boxes with the prompt as text, and a grey shaded background; label outgoing arcs with recognitions
• Prompt states where no recognition occurs:
  – Use rectangular boxes with the prompt as text, and no background
• Application processing steps:
  – Use round-edged boxes with processing specified inside
• Application (non-recognition) decision points:
  – Use diamonds with the conditions specified inside
Welcome to the mood change test.

Are you happy or sad?

Happy = Happy + 1
I'm glad you're happy.

Sad = Sad + 1
I'm sorry you're sad.

Happy > 3

Ok, that's enough happiness. Bye!
Hiding Complexity

Happy = 0; Sad = 0

Welcome to the mood change test.

Happiness
Incrementer
SubDialog

Ok, that's enough happiness. Bye!
Happiness Incrementer Subdialog

Are you happy or sad?

happy

Happy = Happy + 1

I'm glad you're happy.

no

Sad = Sad + 1

I'm sorry you're sad.

yes

Happy > 3

Now think of pizza and tell me:
Form of Submissions

• A PowerPoint file, ‘Save As …’ no later than Office 2003
• First slide has your name, your student ID and a title, typically something like ‘Tutorial 8: Travel Planner Dialogs’
• Keep each call-flow slide relatively simple – use subdialogs
Next Week’s Tutorial

• Everyone is expected to make a 10-15 minute presentation
• There is no Week 9 practical: spend the time on your tutorial presentation
The Tutorial Task

Consider five application types:

- Auto attendants
- Stock quotes and stock trading
- Flight information and flight reservations
- Email access
- Calendar managers
The Tutorial Task

For each application:

• Produce a sample dialog that you will present at the tutorial
• Produce a call-flow diagram for the application
• Assess the application complexity on each of the dimensions we have discussed; rank each app on a scale of 1 (least complex) to 5 (most complex), and present your rankings in a table
• Be prepared to defend your rankings
## The Tutorial Task

<table>
<thead>
<tr>
<th></th>
<th>Vocab</th>
<th>Dialog</th>
<th>Initiative</th>
<th>Syntax</th>
<th>Personalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stocks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calendar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For Next Week’s Class

• Read:

• Be prepared to answer questions about this.