Dialogue systems

- Speak to anyone anywhere

Speech-to-speech translations:

- Robots
- Toys
- Game characters

Provide personalities:

- Answer questions
- Search the web
- Read your email to you

Information navigators:

- News
- Driving directions
- Stocks, weather, flight information

Using speech technology
Dialog Management Internals

- Parser:
  - from recognized words to structure

- Dialog manager:
  - state of dialog
  - references, user profile etc
  - Interface to information

- Language generation:
  - from structure to words
- VIA CLEVELAND: UNSPECIFIED
- AIRLINE: UNSPECIFIED
- TIME: before 12
- DATE: 20001206
- DESTINATION: Boston

Parse it and build semantic frame:

"I wanna go to Boston tomorrow some time in the morning"

from recognized words to structure

Parsing
If no TIME specified, ask

- Keep track of the state of the dialog

- Play adverts

- Stall the user while response comes back

- Deal with bare in, hang ups etc

- Sent to traveley.co.uk

- Build query for database

Dialog management
Should generate marked up text to synthesize morning:

- "US Airways has a flight departing at 07:45 tomorrow.
- "I have US 0775 leaving at 07:45 tomorrow."

Can generate different surface text:

- FLIGHT: US0775
- TIME: 07:45
- DATE: 20001206
- DESTINATION: Boston
- START: Pittsburgh

Build sentence from frame:

Language Generation

From frame structures to (marked up) text
State-based dialog systems
VoiceXML

VoiceXML consortium to build a standard speech interfaces

XML-based markup language

Output synthesized speech and audio files

Recognition of speech and DTMF

Recording of spoken input

Telephone features
“e.g., AT&T’s "How May I help you."
- but only N actions available
- user can say what they like:
  Classification: □
  - system can ask for information
  - user can go where they like
  Mixed Initiative: □
  - can "know" what will be said
  - can switch language models
  - form Hilting paradigm
  System Initiative: □

Dialog types
Most reliable, but many turns

- Fixed dialog structure
- Limited types of interaction

Disadvantages:
- Easy to build
- Harder for the system to be confused
- Harder for the user to be confused
- You know what the user will say (sort of)

Advantages:
- What city are you going to?
- What city are you leaving from?

Simple form filling:

Most existing telephone-based systems

System initiative
Mixed initiative

- can miss important parts
- can get confusing

Disadvantages:

- Can do more complex tasks
- more realistic dialogue

Advantages:

- "Jump" through different parts of dialogue state space
- more interesting dialogue

User and Systems can take initiative:

e.g. Communicator
---

**Classification Dialogs**

---
Building dialog systems

- command and control vs interaction
- how important is natural language
- a document browser or table/search engine
- can you navigate through generic data

Custom vs generic speech interfaces:

- "spontaneous" modality user behavior to make it easier
- CMU’s USI project

Standardized conventions:

-
Building Dialogue Systems

- task completion
- number of repeat calls
- number of calls

How do you know if it's working:
- not needed by experienced users
- "chatty" helpful comments

How do you train them:
- Are users always first time users
- Resources can get very big

How many channels: □
Speaker: Recognizer + Translator + Synthesizer.

Listener: What about integrating these
- What about dialog management
- Just getting this to work isn't easy.
three systems compounding errors
really natural speech is difficult
turn-taking might not translate

Harder things:

pick it up (don’t need to resolve “it”)

don’t need to understand

Easier things:

Should nms and er be translated?

Conversation not sentences

must be fast for dialogue

It’s different from text:

Speech translation
No dialog model (isolated sentences)

- For translation in the field (US Army Chaplains)
- "Rapid deployment in new languages"

Diplomats/Tongues:

- Tourist information domain
- Different partners provide different engines
- 16 partners, 11 languages

CSTAR:

- Some have fixed set of phrases
- Has backup text/voice selection
- Works for simple limited domains

Speech translation today
command control vs engaging dialogue

system vs mixed initiative

Different complexities in dialogue systems:

- Language generation
- Dialogue control
- DB access
- Parsing

Dialogue component:

- Need to control state too
- Recognition + Dialogue + Synthesis

Not just combination of:

Summary