Multi-Modal User Interaction
Fall 2008

Lecture 5: System design and applications

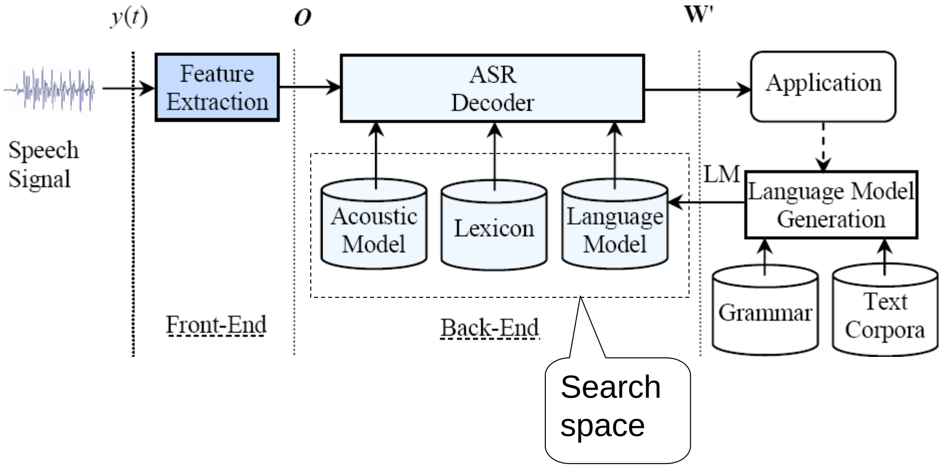
Zheng-Hua Tan

Department of Electronic Systems
Aalborg University, Denmark
zt@es.aau.dk

Part I: Designing HMM-based ASR systems

- Designing HMM-based ASR systems
  - Isolated word recognition
  - Word sequence recognition
  - Optimal graph structures for language decoding
- Applications

Speech recognition system



Designing HMM-based ASR systems

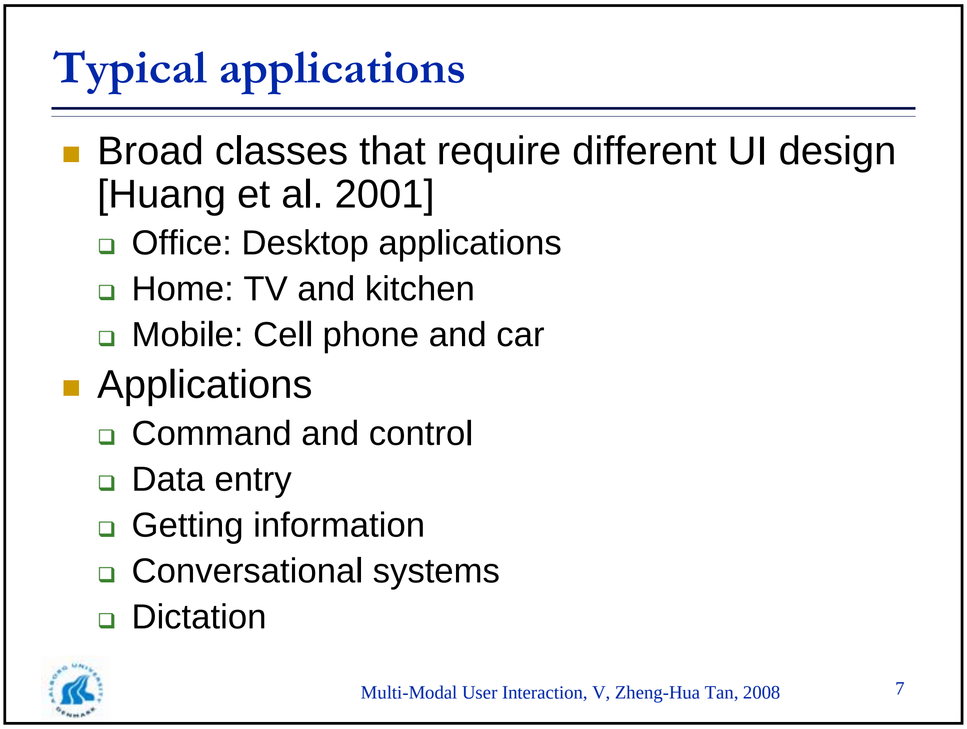
- Lecture Notes at MIT by Rita Singh (Guest lecturer).
  - Language models, acoustic models and search (decoding)

Part II: Applications

- Designing HMM-based ASR systems
- Applications
  - Command and control
  - Telephony applications
  - Dictation

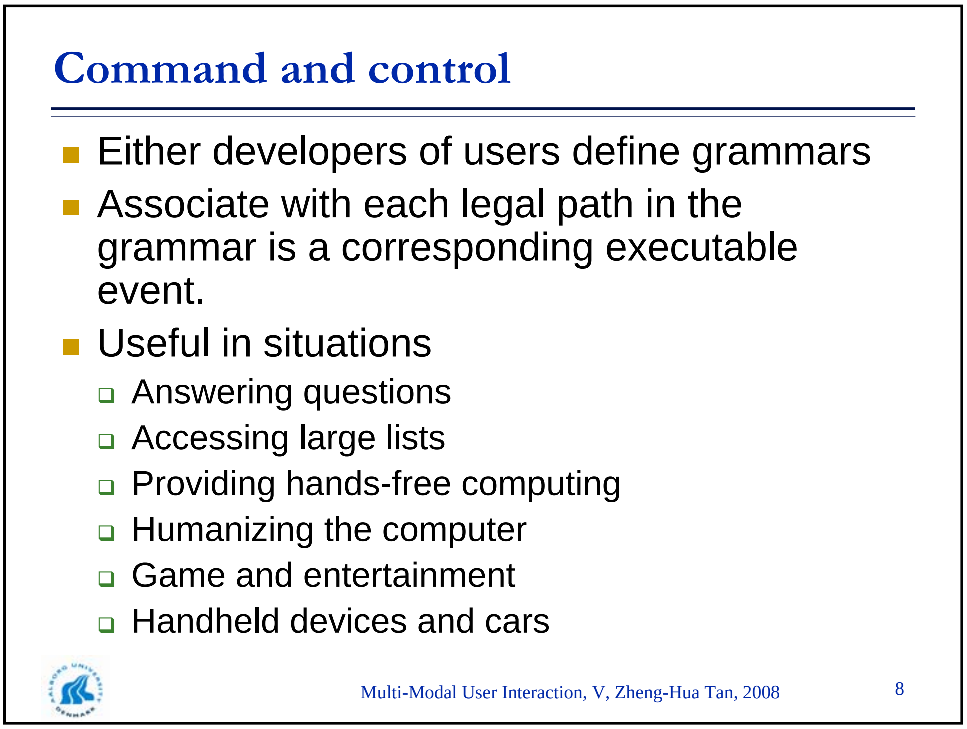
Attributes of ASR systems

- Vocabulary: small (<20 words) to large (>50K words)
- Perplexity: small (< 10) to large (> 200)
- Enrollment: speaker-dependent to speaker-independent
- Speaking mode: isolated-word to continuous-speech
- Speaking style: read speech to spontaneous speech
- SNR: high (> 30 dB) to low (< 10 dB)
- Transducer: noise-concelling microphone to cell phone



Typical applications

- Broad classes that require different UI design [Huang et al. 2001]
  - Office: Desktop applications
  - Home: TV and kitchen
  - Mobile: Cell phone and car
- Applications
  - Command and control
  - Data entry
  - Getting information
  - Conversational systems
  - Dictation



Command and control

- Either developers of users define grammars
- Associate with each legal path in the grammar is a corresponding executable event.
- Useful in situations
  - Answering questions
  - Accessing large lists
  - Providing hands-free computing
  - Humanizing the computer
  - Game and entertainment
  - Handheld devices and cars

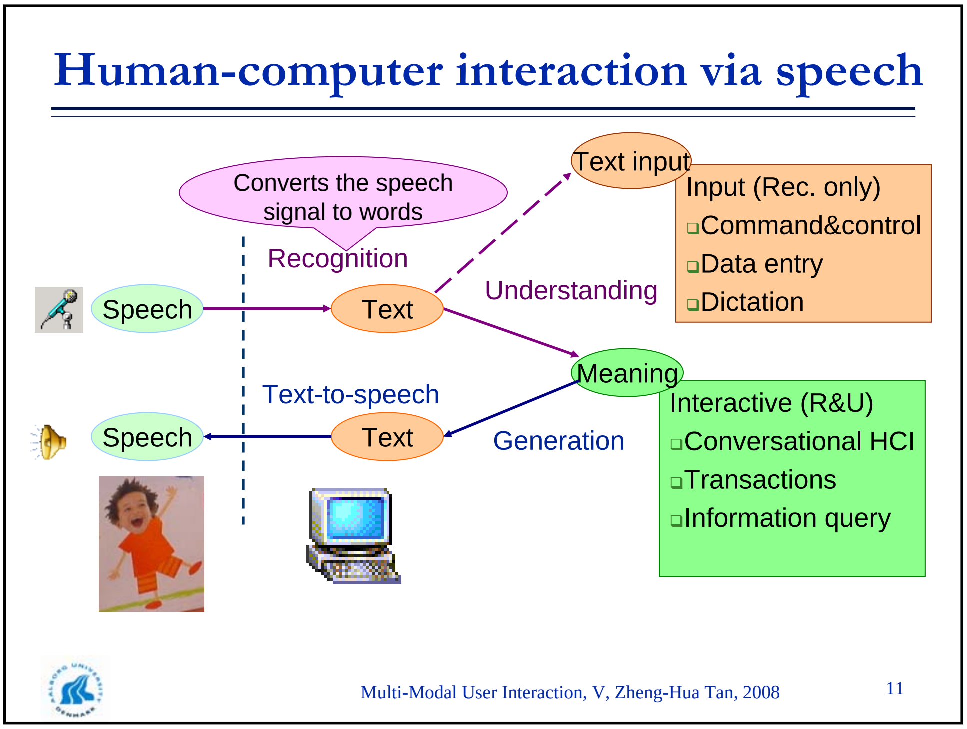
Telephony applications

- Dialogue Example of the Philips Train-timetable information system deployed in 1995 and is used daily (call +49241604020) [from Lars Bo Larsen]:

1. System: Good morning. This is the automatic timetable information service. How can I help you?
2. User: I need a connection from Hamburg to Munich.
3. System: When would you like to go from Hamburg to Munich?
4. User: Next Friday at eight o’clock.
5. System: So you want to travel on Sunday the 20th of November at eight o’clock a.m.?
6. User: No, on Friday.
7. System: So you want to travel on Friday the 18th of November at eight o’clock a.m.?
8. User: Yes.
9. System: There is the following connection: With Intercity Express 793, departure from Hamburg at 8:02, arrival in Munich at 14:05. Do you want me to repeat the connection?

Dictation

- Dictation should not be considered “general recognition”, as it is dependent on the ”topic” of the text data used for LM-training
- Dictation performs better after adaptation to the user
  - Though it can be used as speaker-independent.



Human-computer interaction via speech

Converts the speech signal to words

Recognition

Text input

Input (Rec. only)
- □Command&control
- □Data entry
- □Dictation

Speech

Text

Understanding

Meaning

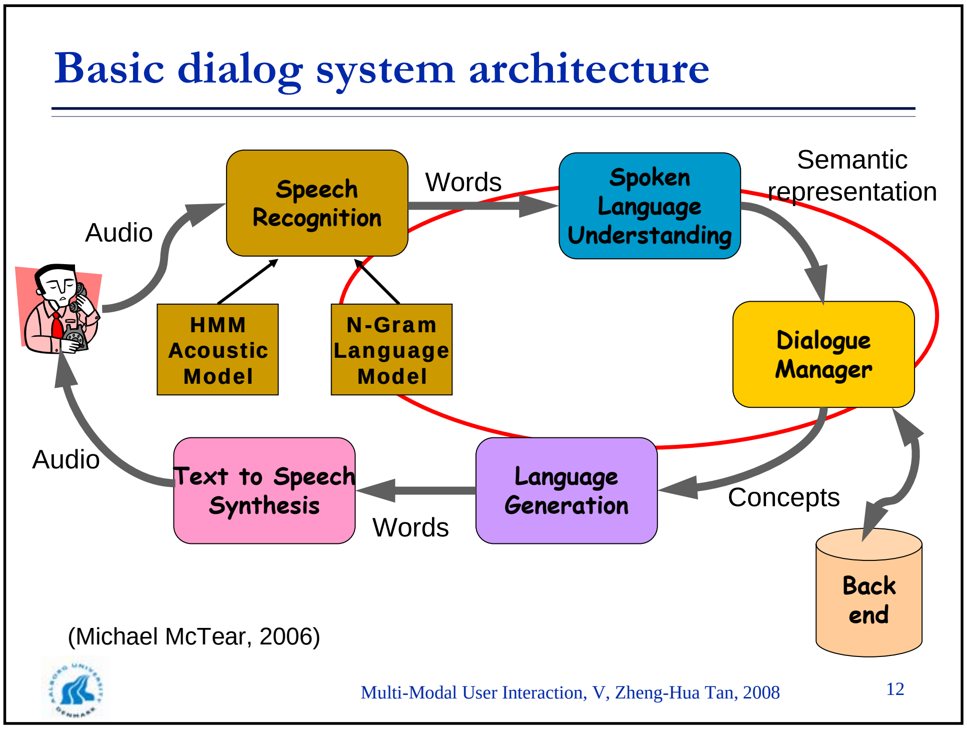
Text-to-speech

Speech

Text

Generation

Interactive (R&U)
- □Conversational HCI
- □Transactions
- □Information query



Basic dialog system architecture

Speech Recognition

Words

Spoken Language Understanding

Semantic representation

Audio

HMM Acoustic Model

N-Gram Language Model

Dialogue Manager

Audio

Text to Speech Synthesis

Language Generation

Concepts

Words

Back end

(Michael McTear, 2006)

Kitchen scenario – fact or fiction?

- Rachel goes into the kitchen, takes a piece of bread and puts it into the toaster. “Not so well done this time.” She goes to the fridge, takes out a carton of milk, and notices that it is almost empty. “Don’t forget to order another carton of milk”, she says to the fridge. “You’re having some friends round for hot chocolate later, maybe I should order two cartons”, says the fridge. “Okay”, says Rachel.

  (McTear)

Summary

- Designing HMM-based ASR systems
  - Isolated word recognition
  - Word sequence recognition
  - Optimal graph structures for language decoding
- Applications
  - Command and control
  - Telephony applications
  - Dictation