Fundamentals of Spoken Language Processing

CECS401

Note-1

Tuesday 8/24/99
Office hours: Wed, 3:30-5:30 PM
E-mail: zhao@cecs.missouri.edu
Fax: 882-8318
Phone: 882-3374
Office: RM 335, EBW
Yunxin Zhao

Instructor:

CECS401: Fundamentals of Spoken Language Processing
Lecture notes
Textbook

Text

Hand Project
Literature Review
Exam

Homework and Lab assignments

Expected Work

Experience with C Programming and UNIX

Probability Theory, Statistics

Signal Processing

Expected background knowledge
The course will be focused on statistical models of speech and language, including hidden Markov models, EM algorithm, and a number of newly innovated statistical learning algorithms.

- speech production
- speech perception
- speech analysis
- speech synthesis
- speech coding
- language modeling
- language recognition
- other speech processing topics
Applications of Spoken Language Technology

- Speech compression in wireless and wireless communications
  and digital libraries
- Wideband speech and audio compression for teleconferencing
- Multimedia communications
  - Keyboard-free multimedia computer devices
  - Voice control in immersive virtual-reality environment and
    social qualities
  - Interactive agent that can talk and listen with human and
    mouse, etc
- Input modalities: speech, pen, hand-gesture, touch, keyboard
- Multimodal human-computer interface
Fig. 1.1: Block-diagram of a typical speech recording system (sampling).

Acquisition

A. Speech signal

Topic 1: Introduction to speech processing and recognition.
Information: 6 words, 22 phonemes, 9 syllables
Duration: 2.02 seconds
Sampling rate: 10.67 KHz
Sentence: Bright sunshine shimmers on the ocean
Example
Among early efforts in continuous speech recognition, this example is taken from R. Reddy's work in late 60's, which

Sentence: Bishop moves to King Knight five
Figure 1.1
The sentence "Bishop moves to king knight five" aligned with its speech waveform
optimal way.

partially or in lieu of these knowledge sources in a sufficiently
hidden Markov model (HMM) of speech provides a powerful
statistical model of communication theory. In particular,
acoustic-phonetic, linguistic, speech recognition, speech processing,
acoustic processor and linguistic decoder are based upon

The acoustic processor and linguistic decoder are based upon

Source-channel model of speech recognition

B. State-of-the-art
necessary.

Speech recognition systems can work well under constrained environments for constrained tasks, but are far from being capable of accurate recognition of speech in arbitrary discourse (may not be). In the past decade, speech recognition techniques have made significant progress.
matches, mismatched training-retest conditions.
Channel distortion, background noise, types of transducer.
Acoustic environment:
Level of perplexity (word-branchnine task).
Syntax:
Small, medium, large, close, open.
Vocabulary:
Speech
Isolated words, continuous speech, read speech, conversational.
Speaking style:
Gender, dialect, accent, age, emotion, speech rate, etc.
Speaker:
Constraints: