CECS401
Fundamentals of Spoken Language Processing

Note-1

Tuesday 8/24/99

CECS401: Fundamentals of Spoken Language Processing

Instructor:

Yunxin Zhao
Office: Rm 335, EBW
Phone: 882–3374
Fax: 882–8318
Email: zhao@cecs.missouri.edu
Office hours: Wed. 3:30 -5:30 PM

Expected background knowledge

Signal processing
Probability theory, statistics
Experience with C programming and UNIX

Expected work

homework and lab assignments
exam
literature review
final project

Text

textbook
lecture notes

Topics

- speech production
- speech perception
- speech analysis
- speech coding
- speech synthesis
- speech recognition
- language modeling
- other speech processing topics

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.

Applications of spoken language technology

- Multimodal human-computer interface
  - —Input modalities: speech, pen, hand-gesture, touch, keyboard, mouse, etc
  - —Interactive agent that can talk and listen with human and social qualities.
  - —Voice control in immersive virtual-reality environment and in hands-busy eyes-busy conditions
  - —Keyboard-free miniaturized computer devices.
- Multimedia communications
  - —Wideband speech and audio compression for teleconferences and digital libraries.
  - —Speech compression in wireline and wireless communications.

- Security
  - —Talker identification and verification
- Aids to handicapped people
  - —Dictation machine
  - —Hearing aids

Topic-1. Introduction to speech processing and recognition

A. Speech signal

Acquisition

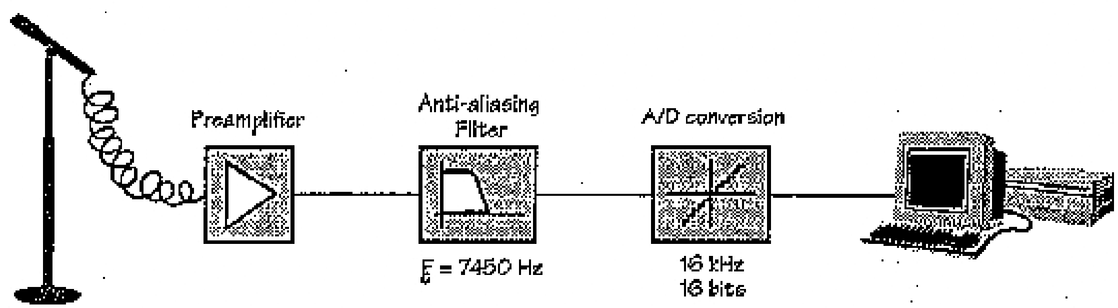


Fig. 1.1. Block-diagram of a typical speech recording system (sampling frequency = 16kHz in this case).

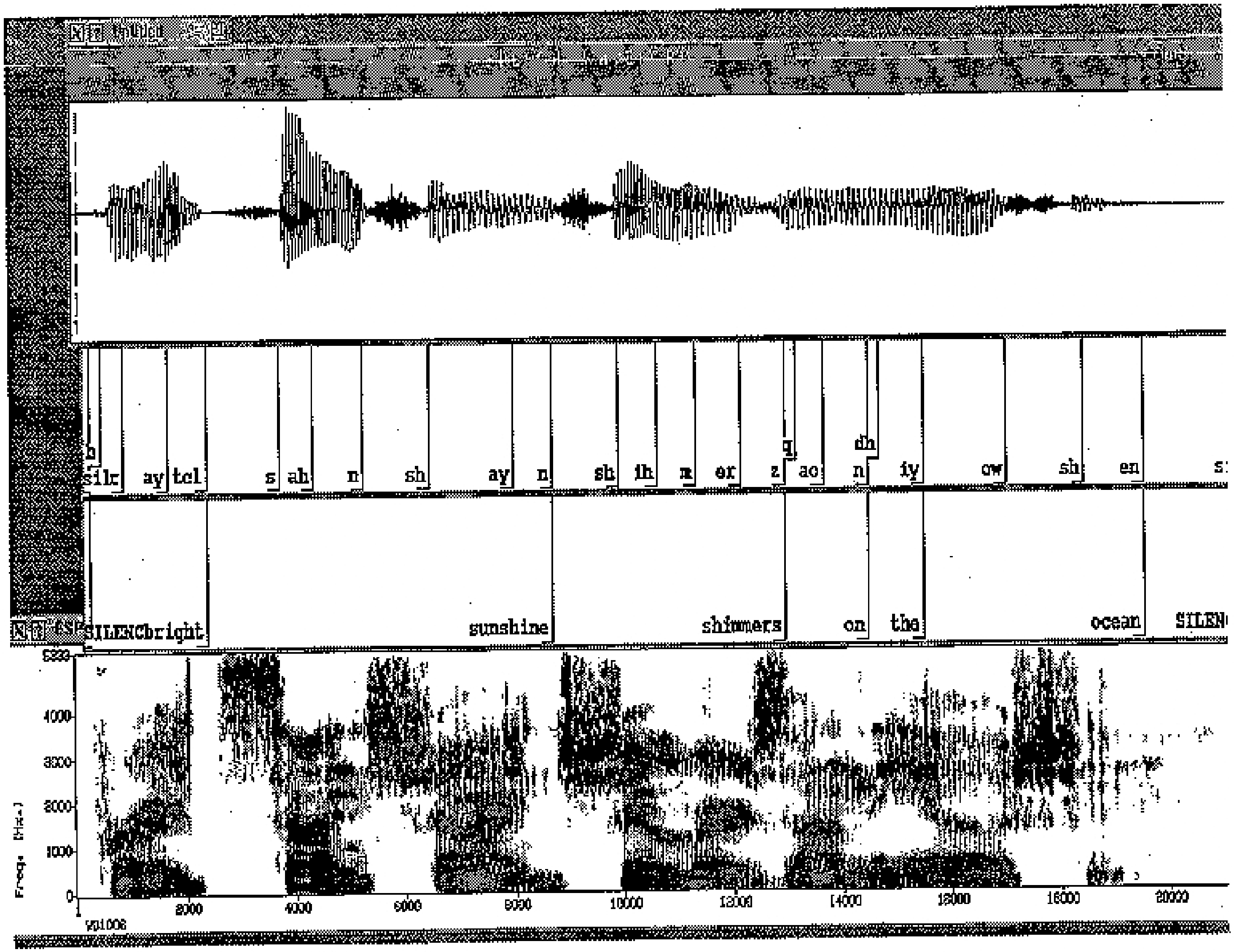
Example

Sentence: Bright sunshine shimmers on the ocean

Sampling rate: 10.67 KHz

Duration: 2.02 seconds

Information: 6 words, 22 phonems, 9 syllables



h
sil
r
ay
tcl
s
ah
n
sh
ay
n
sh
ih
m
er
z
q
ao
dh
n
iy
ow
sh
en
SILENC
bright
sunshine
shimmers
on
the
ocean
SILEN
Freq: (Hz)
5333
4000
3000
2000
1000
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000

Example

Sentence: Bishop moves to king knight five

This example is taken from Raj Reddy's work in late 60's, which was among early efforts in continuous speech recognition.

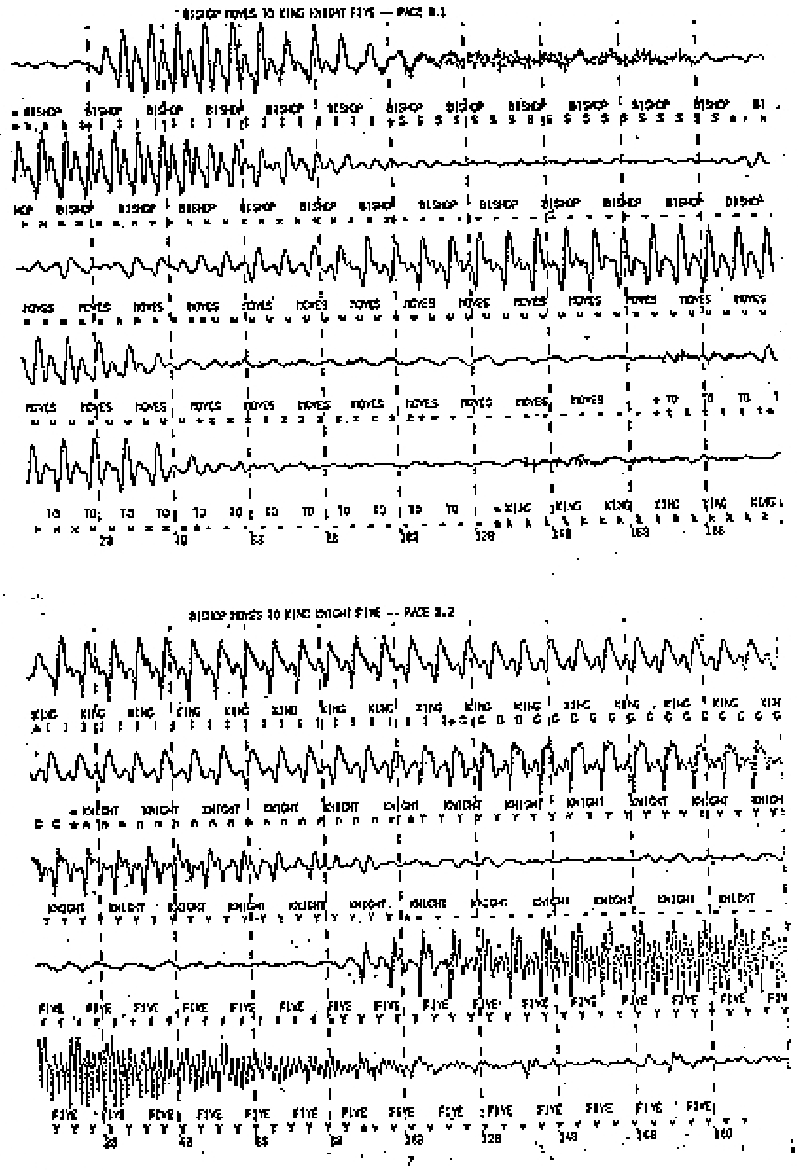
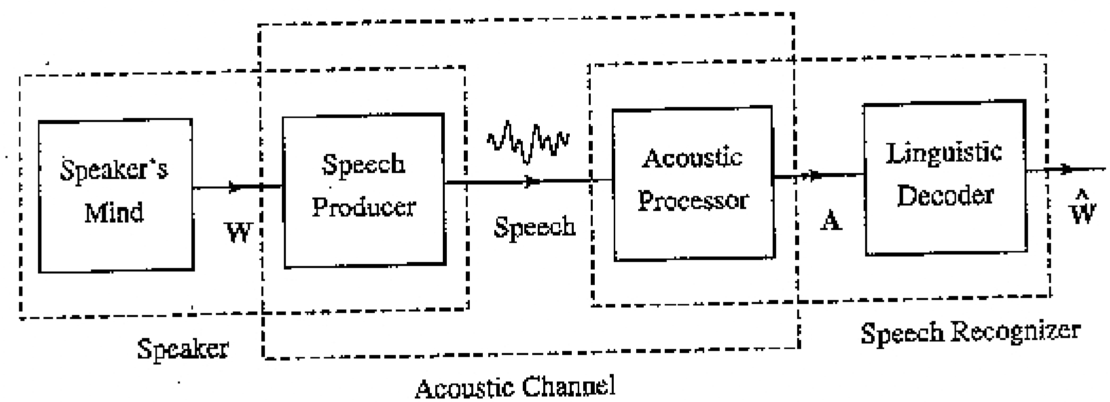


Figure 1.1
The sentence "Bishop moves to king knight five" aligned with its speech waveform

B. State-of-the art

Source-channel model of speech recognition



The acousitc processor and Linguistic decoder are based upon acoustic-phonetics, liguistics, speech perception, signal processing, statistical modeling, and communication theory. In particular, hidden Markov modeling (HMM) of speech provides a powerful paradigm of integrating these knowledge sources in a statistically optimal way.

Speech recognition techniques have made significant progresses in the past decade.

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 necessary).

Constraints:

Speaker:

Gender, dialect accent, age, emotion, speech rate, etc

Speaking style:

Isolated words, continuous speech, read speech, conversational speech

Vocabulary:

Small, medium, large, close, open

Syntax:

level of perplexity (word-branching factor)

Acoustic environment:

Channel distortion, background noise, types of transducer, matched, mismatched training-test conditions.