

Automatic Speech Recognition (ASR)

Steve Renals & Hiroshi Shimodaira
Centre for Speech Technology Research (CSTR)

Semester: 2 (12/Jan/2009 – 27/Mar/2009)
Credit Level: 11
Credit Points: 10
Pre-requisite course: SP (PPLS)
Date & Time: Monday, Thursday, 15:00-15:50
Room: AT 4.12
Course web page:
<http://www.inf.ed.ac.uk/teaching/courses/asr/>

Why learning ASR is important ?

- SR is one of the typical human intelligent activities.
ASR: finds practical ways of how machines can imitate the activities
 - Tackling real world problems ↔ toy problems
↓
complex / mixed research domains
 - audio signal processing
 - pattern recognition / machine learning
 - natural language processing
 - speech understanding
 - dialogue control / planning
 - system integration
- key idea: **statistical approach, optimisation**

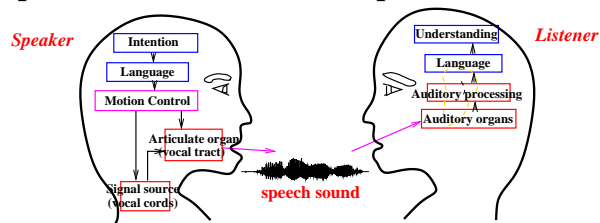
ASR 2008-9 (H. Shimodaira)

2

What is ASR ?

Research on speech processing:

- **Science** Reveal mechanisms or rules how humans
 - generate speech
 - recognise speech
 - communicate each other with the help of speech
- **Engineering** develop systems that
 - synthesise speech like real humans
 - **recognise / understand humans speech like humans do**
 - provide natural human-computer interface



ASR 2008-9 (H. Shimodaira)

1

Why learning ASR is important ? (cont. 2)

- **Impacts**
 - signal processing: LPC, PARCOR
 - pattern recognition: DP
 - machine learning: HMM, GMM
 - language processing: statistical LM (n -gram)

ASR 2008-9 (H. Shimodaira)

3

Course description

SP (PPLS) : intuitive introduction to ASR + TTS
← **pre-requisite courses**

ASR (Informatics) : practical ASR
more theories, algorithms and maths !
+
experiments (developing an ASR system)

■ Recommended courses: "Learning from Data" (LFD), "Probabilistic Modelling and Reasoning" (PMR), "Empirical Methods in Natural Language Processing" (EMNLP),

■ Assessment weighting:

Written examination: 70%

Assessed assignments: 30% (one assignment)

Reference Texts

■ for basic understanding:

■ John N. Holmes, Wendy J. Holmes, "Speech Synthesis and Recognition", Taylor and Francis (2001), 2nd edition

■ for deep understanding:

■ Xuedong Huang, Alex Acero and Hsiao-Wuen Hon, "Spoken language processing: a guide to theory, algorithm, and system development", Prentice Hall (2001).

■ Lawrence R. Rabiner and Biing-Hwang Juang, "Fundamental of Speech Recognition", Prentice Hall (1993).

■ B. Gold, N. Morgan, "Speech and Audio Signal Processing: Processing and Perception of Speech and Music", John Wiley and Sons (1999).

Schedule

Week		Monday		Thursday
1	12.Jan	Signal analysis	15.Jan	Signal analysis
2	19.Jan	Statistical PR	22.Jan	LAB
3	26.Jan	Statistical PR	29.Jan	LAB
4	02.Feb	HMM	05.Feb	HMM
5	09.Feb	Context-dependent model 1	12.Feb	Context-dependent model 2
6	16.Feb	Lexicon and language model	19.Feb	LAB
7	23.Feb	Search and decoding	26.Feb	LAB
8	02.Mar	Speaker adaptation 1	05.Mar	LAB
9	09.Mar	Speaker adaptation 2	12.Mar	Case study
10	16.Mar	Current topics	19.Mar	Review
11	23.Mar	(no lecture)	26.Mar	(no lecture)

Week 1 – 3 & Labs

H.Shimodaira

Week 4–11

S.Renals

Before start ...

■ There are a lot of maths!

No need to understand them in details, but try to get ideas.

■ Lectures (as in syllabus) + Lab sessions (exercises)

■ Lab session:

- Analysis of speech signals
- Generative models (GMM and EM-algorithm)
- Continuous speech recognition
- Language models
- Continuous speech recognition system

■ Contact information

Hiroshi Shimodaira, School of Informatics

Email: h.shimodaira@ed.ac.uk Tel. ext. 51 3279

Office: IF 3.04, CSTR, 10 Crichton Street