

MPhil in Advanced Computer Science

Advanced Speech Technology

Lecturers: Phil Woodland, Blaise Thomson, Paul Taylor
Timing: Lent Term
Prerequisites: Module L106 “Spoken Language Processing” or equivalent experience
Structure: 16 lectures

AIMS

The aim of this module is to describe techniques used in, and architectures for, the design of state-of-the-art speech technology systems. These methods are starting to appear in many types of information processing and computer systems. The course focusses on three main areas: speech recognition; spoken dialogue systems and text-to-speech speech synthesis.

SYLLABUS

The syllabus is divided into three main parts: advanced techniques for speech recognition (8L); spoken dialogue systems (4L) and text-to-speech synthesis (4L).

- **Advanced Techniques for Speech Recognition Systems (Prof. Phil Woodland)**
 1. Introduction and review of speech recognition components
 2. Large vocabulary speech recognition search & generation of multiple hypotheses.
 3. Lattices, confusion networks & confidence estimation
 4. Discriminative training
 5. Feature Selection and Extraction: PCA, LDA
 6. Audio segmentation and speaker clustering
 7. Adaptation & Robustness: MAP, linear transforms, MLLR, noise robustness
 8. Speech Recognition system examples and applications
- **Spoken Dialogue Systems (Dr. Blaise Thomson)**
 1. Introduction to SDS: Requirements; basic grammar-driven systems, confidence measures and N-Best hypotheses, and state-of-the-art.
 2. Semantic Processing: Approaches to semantic processing, robust parsing, template matching, statistical decoding, The flat concept model, the hierarchical model.
 3. Dialogue Act Decoding: Decision trees, belief networks, latent semantic analysis.
 4. Dialogue Management: Flow-chart methods, information update approach, statistical dialog management, Markov Decision Processes, the Hidden Information State model.
- **Text to Speech Synthesis (Dr. Paul Taylor)**
 1. Overview of text-to-speech. Text analysis.
 2. Pronunciation: Lexicon vs rules, phonesets, modelling variation in pronunciation, regional accents. Pronunciation for unknown words; letter-to-sound rules. Stress and accentuation.
 3. Basic Speech Generation Techniques: Formant synthesis, basic diphone synthesis; signal processing for synthesis. Intonation and F0 control.
 4. Unit Selection: Data-driven synthesis; unit selection paradigm; target costs, join costs, search. Automatic voice enrolment; future directions.

OBJECTIVES

On completion of the module students should have

- Understand advanced techniques used for building speech recognition systems
- Understand the components, issues and approaches for constructing spoken dialogue systems
- Understand techniques and approaches for text-to-speech synthesis
- Be aware of the current state-of-the-art in each of the areas covered

ASSESSMENT

The assessment of the module will be by two essays chosen from a set of suggested topics: one essay will be in the area of speech recognition and the second in the area of either spoken dialogue systems or text-to-speech synthesis. Each essay will contribute to 50% of the marks.

RECOMMENDED READING

- Daniel Jurafsky and James Martin. *Speech and Language Processing*, 2nd edition, Prentice- Hall 2008
- Xuedong Huang, Alex Acero and Hsiao-Wuen Hon. *Spoken Language Processing*, Prentice- Hall, May 2001.
- Paul Taylor, *Text-to-Speech Synthesis*, Cambridge University Press, February 2009