

MPhil in Advanced Computer Science

Spoken Language Processing

Lecturers:	Mark Gales, Adria de Gispert, Phil Woodland
Timing:	Michaelmas term
Prerequisites:	None
Structure:	14 lectures + 2 practical sessions

AIMS

The aim of this module is to introduce the underlying statistical approaches and some of the major techniques used for spoken language processing. Core statistical models that are used in a wide-range of speech and language applications will be discussed along with their underlying theory. Examples of how these models may be applied to speech processing applications, such as speech recognition and speaker verification, will be described.

SYLLABUS

- **Introduction**
Scope of the course and introduction to speech sounds and frequency domain representations.
- **Basic Pattern Processing**
Bayes' decision rule, forms of statistical classifier and generative models.
- **GMM-based Speaker Verification**
Basic feature vectors (MFCC/deltas), MAP parameter estimation, decision rules/ROC curve.
- **Hidden Markov Models and the Viterbi algorithm**
HMM structure and underlying assumptions, training using Baum-Welch and the EM algorithm. Networks and the Viterbi algorithm.
- **Decision Trees and Context Modelling**
Phone-level variation, dictionaries, decision trees and context clustering.
- **Weighted Finite State Transducers**
Basic operations of WFST, and WFST representation of information in speech systems.
- **N-gram Language Models**
N-gram language models, discounting, smoothing, backing-off, mixture language models and interpolation. WFST representation of language models.
- **Applications of spoken language processing**
Examples of applications, including speech recognition and speech synthesis.

OBJECTIVES

On completion of the module students should:

- understand the basic principles of pattern classification;
- understand Gaussian mixture models, hidden Markov models and N-gram language models;
- understand and be able to implement the Viterbi algorithm;
- understand weight finite state transducers;
- be able to apply the above approaches to speech processing applications.

PRACTICAL WORK

- Practical 1: Speech Recognition System: implementation of Viterbi algorithm and design of a basic continuous speech recognition system.
- Practical 2: Language Modelling: implement language model interpolation and language model performance evaluation.

ASSESSMENT

The module will be assessed by two practical reports. Each report will contribute 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.
- Richard Duda, Peter Hart and David Stork: *Pattern Classification*, Second Edition, John Wiley & Sons Inc, 2000. ISBN 0471056693