

# ELEC9723

## Speech Processing

THE UNIVERSITY OF  
NEW SOUTH WALES



School of Electrical Engineering  
and TelecommunicationS

### COURSE INTRODUCTION — Session 1, 2009

#### Course Staff

Course convener: Mohaddeseh Nosratighods, [hadis@unsw.edu.au](mailto:hadis@unsw.edu.au)

Laboratory demonstrator: Vidhyasaharan Sethu, [vidhyasaharan@gmail.com](mailto:vidhyasaharan@gmail.com)

#### Course details

**Credits:** The course is a 6 UoC course; expected workload is 9-10 hours per week throughout the 12 week session.

**Contact hours:** The course consists of 3 hours of per week, comprising lectures and/or laboratory (a typical class might be 1½ hours of lecture followed by 1½ hours of lab):

Lectures: Tuesdays, 6pm–9pm, room EEG3

Lab sessions: Tuesdays, 6pm-9pm, room EE214

**Consultations:** You are encouraged to ask questions on the course material before the regular class times (e.g. from 5:45pm) at the lecture venue in the first instance, rather than by email.

#### Course Information

##### Context and aims

ELEC9723 Speech Processing builds directly on students' skills and knowledge in digital signal processing gained during ELEC3104 Signal Processing and ELEC4621 Advanced Digital Signal Processing. Speech processing has been one of the main application areas of digital signal processing for several decades now, and as new technologies like voice over IP, automated call centres, voice browsing and biometrics find commercial markets, speech seems set to drive a range of new digital signal processing techniques for some time to come. This course provides not only the technical details of ubiquitous techniques like linear predictive coding, Mel frequency cepstral coefficients, Gaussian mixture models and hidden Markov models, but the rationale behind their application to speech and an understanding of speech as a signal. Contemporary signal processing is almost entirely digital, hence only discrete-time theory is presented in this course.

**Aims:** This course aims to:

- a. Familiarise you with modeling the vocal tract as a digital, linear time-invariant system, and the extraction of parameters that describe it.
- b. Convey details of a range of commonly used speech feature extraction techniques.

- c. Provide a basic understanding of multidimensional techniques for speech representation and classification methods.
- d. Familiarise you with the practical aspects of speech processing, including robustness, and applications of speech processing, including speech enhancement, speaker recognition and speech recognition.
- e. Give you practical experience with the implementation of several components of speech processing systems.

### **Relation to other courses**

ELEC9723 Speech Processing is the most advanced course offered by the university on this topic, and serves as an excellent basis from which to commence research in the area. Various aspects of the course bring students up to date with the very latest developments in the field, as seen in recent international conferences and journals, and some of the laboratory work is designed in the style of an empirical research investigation. ELEC9723 is well complemented by ELEC9724 Audio and Electroacoustics, which deals with many other signal processing methods and gives an understanding of human auditory perception (also a key part of speech processing), discusses compression techniques (many related to speech coding) and an understanding of audio signals. ELEC9723 is also well complemented by ELEC9722 Digital Image Processing, which gives an insight into two-dimensional signal processing and image signals. ELEC9721 Digital Signal Processing Theory and Applications provides an excellent basis for Speech Processing, however for students who have not already completed this course (or ELEC4621), it is recommended for future study.

**Pre-requisites:** The minimum pre-requisite for the course is ELEC3104, Signal Processing (or equivalent). Knowledge from either ELEC4621 or ELEC9721 is highly desirable.

**Assumed knowledge:** It is essential that you are familiar with the sampling theorem, digital filter design, the discrete Fourier transform, stochastic processing. If you are not confident in the knowledge from previous signal processing courses (especially the topics mentioned), you are strongly advised to revise previous course materials as quickly as possible to avoid difficulties in this course.

**Previous course code:** The course replaces previous course ELEC9344 Speech and Audio Processing.

### **Learning outcomes**

On successful completion you should be able to:

1. Describe the key aspects of typical speech signals
2. Express the speech signal in terms of its time domain and frequency domain representations and the different ways in which it can be modelled;
3. Derive expressions for simple features used in speech classification applications;

4. Explain the operation of example algorithms covered in lectures, and discuss the effects of varying parameter values within these;
5. Synthesise block diagrams for speech applications, explain the purpose of the various blocks, and describe in detail algorithms that could be used to implement them;
6. Implement selected components of speech processing systems, including speech recognition and speaker recognition, in MATLAB.
7. Deduce the behaviour of previously unseen speech processing systems and hypothesise about their merits.

The course delivery methods and course content address a number of core UNSW *graduate attributes*; these include:

- a. The capacity for analytical and critical thinking and for creative problem-solving, which is addressed by the tutorial exercises and laboratory work.
- b. The ability to engage in independent and reflective learning, which is addressed by tutorial exercises together with self-directed study.
- c. The skills of effective communication, which are addressed by the viva-style verbal assessment in the laboratory.
- d. Information literacy, which is addressed by the homework.

Please refer to <http://www.ltu.unsw.edu.au/content/userDocs/GradAttrEng.pdf> for more information about graduate attributes.

## Teaching strategies

The course consists of the following elements: lectures, laboratory work, and home work comprising self-guided study and a problem sheet.

### Lectures

During the lectures, techniques for the analysis, modeling and processing of the digital speech signal will be presented. The lectures provide you with a focus on the core material in the course, together with qualitative, alternative explanations to aid your understanding. Various examples will be given, to enrich the analytical course content. The lectures materials distributed in class (or via the course web site) will give a good guide to the course syllabus, but you will need to supplement them with additional reading, of the recommended text book and/or other materials recommended by the lecturing staff. In particular, you should not assume that attendance at all lectures (even with a glance or two through the notes), on its own, is sufficient to pass the course.

### Laboratory work

The lecture schedule is deliberately designed to gain practical, hands-on exposure to the concepts conveyed in lectures soon after they are conveyed in class. Generally there will be around one week between the introduction of a topic in lectures and a laboratory exercise on the same topic, sufficient time in which to revise the lecture, attempt related problems and prepare for the laboratory. The laboratory work provides you with hands-on design experience and exposure to simulation tools and algorithms used widely in

speech processing. You must be pre-prepared for the laboratory sessions: the laboratory sessions are short, so this is only possible way to complete the given tasks.

Laboratory classes will start in week 1. You will need to bring to the laboratories:

- A USB drive for storing MATLAB script files
- A laboratory notebook for recording your work
- Your lecture notes, laboratory preparation and/or any other relevant course materials

### **Tutorials and Home work**

The lectures can only cover the course material to a certain depth; you must read the textbook(s) and reflect on its content as preparation for the lectures to fully appreciate the course material. Home preparation for laboratory work provides you with the background knowledge you will need. The tutorials aim to provide in-depth quantitative and qualitative understanding of speech processing theory and methods. Together with your attendance at classes, your self-directed reading, completion of problems from the tutorials and reflection on course materials will form the basis of your understanding of this course. There will be discussion time about tutorial solutions in weeks 6 and 10 after the lectures before mid-session and final exams.

### **Assessment**

Laboratory work: 30%  
Mid-session exam: 10%  
Final examination: 60%

**Laboratory work:** Starting in week 2, the laboratory work will be assessed in order to ensure that you are studying and that you understand the course material. The laboratory assessment is conducted live during the lab sessions, so it is essential that you arrive at each lab having revised lecture materials (and attempted problems from the problem sheet) in advance of each laboratory, and having completed any requested preparation for the labs. Without preparation, marks above 50% may be difficult to obtain. No lab reports are required in this course. During the laboratory, you may consult with others in the class, but *you must keep your own notes of the laboratory*. In particular, note that *laboratory assessment will be conducted individually, not on a per-group basis*. Please also note that *you must pass the laboratory component in order to pass the course*.

**Mid-session examination:** The mid-session examination tests your general understanding of the course material, and questions may be drawn from any course material up to the end of week 6. This examination provides you with the feedback on your strengths and weaknesses which assists the learning process and thereby sustains a sense of motivation and interest.

**Final examination:** The exam in this course is a standard closed-book 3 hours written examination, comprising five compulsory questions. University approved calculators are allowed. The examination tests analytical and critical thinking and general understanding of the course material in a controlled fashion. Questions may be drawn from any aspect

of the course, unless specifically indicated otherwise by the lecture staff. Please note that *you must pass the final exam in order to pass the course.*

## Course Schedule

Week	Lecture	Ref	Lecturer	Laboratory
0	No lecture		-	-
1, 10 <sup>th</sup> March	Introduction to speech processing	[1]	Nosratighods	Introductory speech analysis – no assessment
2	Speech analysis	[1]	Nosratighods	Lab 1: Spectral analysis
3	Fundamentals of speech science		Morrison (ANU)	No lab
4	Linear predictive coding	[1,2]	Nosratighods	Lab 2: Feature extraction
5	Time-frequency analysis	[1]	Nosratighods	Lab 3: Linear predictive coding
6	Speech enhancement	[1]	Nosratighods	Lab 4: Speech synthesis using LPC
7	<b>Mid-session examination, duration 1 hour 15 min</b>		Nosratighods	No lab
8	Robust front-end processing		Nosratighods	Lab 5: Front-end processing
9	Clustering and Gaussian mixture models and Speaker Recognition	[1]	Nosratighods	Lab 6: Robust Front-end processing
10	Hidden Markov Model	[1]	Nosratighods	Lab 7: Speaker recognition
11	Speech recognition	[2]	Epps	Lab 8: Speech recognition
12	Speech synthesis		Chen (NICTA)	No lab

## Resources

### Textbooks

#### Prescribed textbook

The following textbook is prescribed for the course:

- [1] Huang, X., Acero, A., and Hon, H.-W. (2001). *Spoken Language Processing*, Prentice-Hall, New Jersey.

Unfortunately there is no single text that covers all topics in the same depth as the course, however the above reference should suffice for the majority of the course. Additional references, listed below and at the end of some lecture note sets, will in combination provide complete coverage of the course. Lecture notes will be provided, however note that these do not treat each topic exhaustively and additional reading is required.

#### Reference texts

The following books are good additional resources for speech processing topics:

- [2] Deller, J. R., Proakis, J. G. and Hansen, J. H. L. (1993). *Discrete-Time Processing of Speech Signals*, Macmillan, Toronto. – Although not as clearly written as other speech processing texts, this is nonetheless a classic reference.
- [3] Quatieri, T. F. (2002). *Discrete-Time Speech Signal Processing*, Prentice-Hall, New Jersey. – Previously this was the recommended text for ELEC9723. It is now out of print, but the library still holds a copy.
- [4] O’Shaughnessy, D. (1987), *Speech Communication: Human and Machine*, Addison-Wesley, Reading, MA. – This is a classic text on speech modeling and processing, which is relevant to many of the early topics.
- [5] Rabiner, L. R., and Juang, B.-H. (1993). *Fundamentals of Speech Recognition*, Prentice-Hall, New Jersey. – This is helpful for the HMM and speech recognition topics.
- [6] Rabiner, L. R., and Schafer, R. W. (1978). *Digital processing of speech signals*, Prentice-Hall, New Jersey. – This is an older text that is helpful mainly for roughly the first half of the course.

### **Books covering assumed knowledge**

The following books cover material which is assumed knowledge for the course:

### **On-line resources**

Some additional on-line resources relevant to the course:

Resource: course webct

<http://vista.elearning.unsw.edu.au>

Library resources

<http://info.library.unsw.edu.au/web/services/teaching.html>

VOICEBOX: Speech Processing Toolbox for MATLAB

<http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>

Speech at Carnegie Mellon University:

<http://www.speech.cs.cmu.edu/>

## **Other Matters**

### **Academic Honesty and Plagiarism**

Plagiarism is the unacknowledged use of other peoples work, including the copying of assignment works and laboratory results from other students. Plagiarism is considered a serious offence by the University and severe penalties may apply. For more information about plagiarism, please refer to <http://www.lc.unsw.edu.au/plagiarism>

### **Continual Course Improvement**

The course is under constant revision in order to improve the learning outcomes of its students. Please forward any feedback (positive or negative) on the course to the course convener or via the Course and Teaching Evaluation and Improvement Process (surveys at the end of the course).

### **Administrative Matters**

On issues and procedures regarding such matters as special needs, equity and diversity, occupational health and safety, enrolment, rights, and general expectations of students, please refer to the School policies, see <http://scoff.ee.unsw.edu.au/>.

### **CATEI Results (S1, 2008)**

The university strongly encourages students to give their feedback at the conclusion of the course. Results from an online survey of ELEC9723 Speech and Audio Processing in 2008 are shown below. In 2009, we will be endeavouring to improve on the quality of the feedback given to you, developing thinking skills, and tutorial support. Please note that the survey assumes that respondents have attended at least 80% of the class contact time.

Q1. This lecturer communicated effectively with students (e.g. He / She explained things clearly). Agree: 100% Disagree: 0%

Q2. This lecturer stimulated my interest in the subject matter he/she was teaching Agree: 100% Disagree: 0%

Q3. This lecturer encouraged me to think critically Agree: 100% Disagree: 0%

Q4. This lecturer provided feedback to help me learn Agree: 66% Disagree: 34%

Q5. This lecturer encouraged student input and participation during classes Agree: 100% Disagree: 0%

Q6. This lecturer was generally helpful to students Agree: 100%

Q7. The lecturer enhanced my ability to learn and apply engineering methods and technology. Agree: 100% Disagree: 0%

Q8. The lecturer explained the links between this subject area and other areas and the profession. Agree: 100% Disagree: 0%

Q9. The lecturer encouraged me to learn independently and develop my own understanding of this subject area. Agree: 100% Disagree: 0%

Q10. Overall, I was satisfied with the quality of this lecturer's teaching Agree: 100% Disagree: 0%