Two courses in one

- Undergraduate course code: LASC10061
- Postgraduate course code: LASC11065

- Shared:
  - Course materials, classes, labs
  - Exam/coursework weightings

- Differences:
  - Undergraduate course: 20 credits
  - Postgraduate course: 10 credits
  - Coursework expectations

Assessment

- Three items of assessment : two practical lab-based assignments and an exam
    1. First assignment: speech synthesis practical **write-up** (20%)
    2. Second assignment: speech recognition practical **write-up** (30%)
    3. Multiple choice **exam** in December (50%)

- Assignment due dates are given in the weekly schedule
- Exam date is set by the University and will be published in due course

Introduction to the course

- learning outcomes
- delivery
- timetable
- background required
- course coverage
- assignments

Introduction to the course

- learning outcomes
- delivery
- timetable
- background required
- course coverage
- assignments

Learning outcomes

- Overview of the components of **speech recognition** and **speech synthesis** systems
- Understand the main concepts and what **each component does**
- Describe a **simple version** of each component
- See what the **difficult problems** are in recognition and synthesis.
- Use tools for **visualising and manipulating speech waveforms**
- **Experiment** with two speech technology systems (Festival & HTK)
- See knowledge and skills from different areas come together in an **interdisciplinary** field

Introduction to the course

- learning outcomes
- delivery
- timetable
- background required
- course coverage
- assignments

Delivery - speech.zone website & Learn

- The website speech.zone contains almost everything you will need
  - **Video material, reading lists, forums, calendar, coursework instructions**
    - you must have an **account** on this site, so that you can post on the forums - make sure you can log in, and email me if you have any trouble

- You only need to use **Learn** to:
  - Sign up for a lab group
  - Optional *Introduction to Unix* material
  - Submit your coursework

Delivery - what I provide

- Scope of the course
  - the **topic map**

- The course is divided into a sequence of **modules** that lead you through that map
  - visiting every topic once, sometimes more
  - discovering the connections between topics
  - tackling the material in increasing order of difficulty

	THEORY					APPLICATION					
	SPEECH			SIGNAL PROCESSING	PROBABILISTIC MODELLING	SPEECH SYNTHESIS		AUTOMATIC SPEECH RECOGNITION			
	SIGNALS	PRODUCTION	PERCEPTION			FRONT END	WAVEFORM GENERATION	FEATURE EXTRACTION	PATTERN MATCHING	HIDDEN MARKOV MODELS	CONNECTED SPEECH
CONCEPTS	TIME DOMAIN	SOUND SOURCE	PITCH	DIGITAL SIGNAL	DESCRIBING DATA	TOKENISATION & NORMALISATION	WAVEFORM CONCATENATION	SERIES EXPANSION	EXEMPLAR	GENERATIVE MODEL OF SEQUENCES	HIERARCHY
	PERIODIC SIGNAL	HARMONICS	COCHLEA	SHORT-TERM ANALYSIS	DISCRETE & CONTINUOUS VARIABLES	PRONUNCIATION	DIPHONE	FEATURES	DISTANCE		SUB-WORD UNIT
	FREQUENCY DOMAIN	VOCAL TRACT RESONANCE & FORMANTS	MEL SCALE	SPECTRAL ENVELOPE	JOINT, CONDITIONAL, BAYES' FORMULA	PROSODY		FEATURE ENGINEERING	SEQUENCE	HIDDEN STATE SEQUENCE	N-GRAMS
MODELS & DATA STRUCTURES	FILTER	RESONANT TUBE	FILTERBANK	IMPULSE TRAIN	GAUSSIAN	FINITE STATE TRANSDUCER		FEATURE VECTOR	SEQUENCE OF FEATURE VECTORS	HIDDEN MARKOV MODEL	
	IMPULSE RESPONSE	SOURCE-FILTER MODEL	PHONEME	PITCH PERIOD	GENERATIVE MODEL	DECISION TREE			GRID	LATTICE	GRAPH
ALGORITHMS & ANALYSIS				FOURIER ANALYSIS	FITTING A GAUSSIAN TO DATA	HANDWRITTEN RULES	OVERLAP-ADD	MFCCS	DYNAMIC PROGRAMMING (DTW)	DYNAMIC PROGRAMMING (VITERBI)	COMPOSITION ("COMPILING")
				CEPSTRAL ANALYSIS	CLASSIFICATION	LEARNING DECISION TREES	TD-PSOLA			BAUM WELCH	APPROXIMATION (PRUNING)

Delivery - what I provide

- The material on speech.zone is divided into **modules**
  - the **video content** provides only the *bare bones* and there are gaps in coverage
  - you need to flesh out the details by taking **full advantage** of other modes of learning
    - readings
    - classes, including activities, quizzes, etc
    - labs
    - forums

Delivery - what you need to do

- Every week you will **watch all videos** and **do essential readings** before the class
- You can then post questions on the forums
  - feel free to ask questions at **any level** - I'll tell you if you go "off topic"
  - try to choose an appropriate forum / topic / thread (I will re-organise as required)
  - usually, the **most basic questions** are the most helpful

- Classes will help you **synthesise** together all the different sources of information and different ways of learning. No single source will be enough on its own.
- To do well in this course, you need to be an **active learner**, not a passive observer
  - *don't worry - this course does not require much "audience participation"*

Laptops are the single most reported **distracter** to **both** users and fellow students.

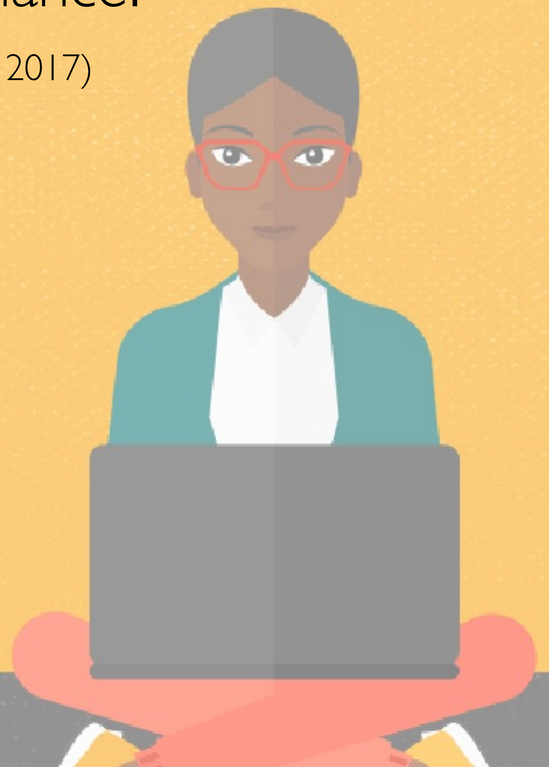
(Fried 2008, Zhu et al 2011)

Laptop & device use is **negatively** correlated with understanding of course material and overall course performance.

(Hembrooke & Gay 2003, Fried 2008, Patterson & Patterson 2017)

The negative effects of device use are the highest among males and low performing students

(Patterson & Patterson 2017)



theguardian : Students who use digital devices in class 'perform worse in exams' by Richard Adams, 11 May 2016.

SCIENTIFIC AMERICAN : Students are Better Off without a Laptop in the Classroom by Cindi May, 11 July 2017.

The Washington Post : Why smart kids shouldn't use laptops in class by Jeff Guo, 16 May 2016.

References

Carrie B. Fried (2008). In-class laptop use and its effects on student learning. **Computers & Education** 50(3): 906-914.

Helene Hembrooke and Geri Gay (2003). The Laptop and the Lecture: The Effects of Multitasking in Learning Environments. **Journal of Computing in Higher Education** 15(1): 46-64.

Richard W. Patterson, Robert M. Patterson (2017). Computers and productivity: Evidence from laptop use in the college classroom. **Economics of Education Review** 57: 66-79.

Erping Zhu, Matthew Kaplan, R. Charles Dershimer, Inger Bergom (2011). Use of Laptops in the Classroom: Research and Best Practices. **University of Michigan CRLT Occasional Paper** 30.

Delivery - class (i.e., lecture) recording

- The video clips on speech.zone are extracted from previous lectures

- Classes are also recorded, ***but…***
  - sometimes the technology fails
  - recordings will not make any sense for interactive parts, activities, quizzes, group exercises, etc…

Delivery - keep making it better

- PDF slides on speech.zone, approximately matching the video clips

- Slides used in classes are improved every year, and will be made available as we proceed
  - these slides are dynamically updated in response to your questions
  - usually, they become available shortly before each class

Delivery - keep making it better

- Please give me **feedback** (email, forum posts, verbally, class reps, PPLS teaching offices, notes slipped under my office door,…) about **course structure** and **delivery mode**, **throughout** the course.

- I also want feedback on speech.zone
  - is it clearly organised?
  - is the website reliable and fast enough?
  - is it obvious what relates to this course, and what does not?
  - does everything work correctly on your device?

Introduction to the course

- learning outcomes
- delivery
- timetable
- background required
- course coverage
- assignments

Timetable

- **Class** - in **two** parts
  - Thursday 09:00 – 09:50
    - type of content varies - see the course calendar
    - for foundation material, you decide whether to attend
  - Thursday 10:00 – 10:50 - always attend
  - Each part of the class will start and finish at the above times (except today)

- **Labs** (you need to attend **one** session every week - sign up for a group on Learn)
  - group 1 : Tuesday 14.10 - 16.00
  - group 2 : Wednesday 11.10 - 13.00

Introduction to the course

- learning outcomes
- delivery
- timetable
- background required
- course coverage
- assignments

What background do you need?

- This course involves:
  - **Linguistics**: phonetics, phonology, prosody
  - **Mathematics**: statistics and probability, optimisation
  - **Engineering**: practical implementations, empirical findings
  - **Computer science**: algorithms, data structures

- So, there will be something new for everyone

What background do you need?

- No single background is assumed, ***but…***
  - If **everything** on that list is **new** to you, then you will probably find the course too hard

- You'll get the most out of this course if you know a little bit about one or more of
  - speech (e.g., phonetics)
  - audio more generally (e.g., music, audio engineering)
  - engineering (e.g., signal processing)
  - computer science

Introduction to the course

- learning outcomes
- delivery
- timetable
- background required
- course coverage
- assignments

Key concepts we will understand

- Simple pattern recognition by comparing to stored examples
- The concept of generative models
  - and how they can be used to make a classifier
- The importance of choosing the right representation of the speech signal
  - feature extraction
  - feature engineering
- Combining prior knowledge and new evidence in a principled probabilistic framework
  - the Bayesian approach
  - modelling variability and uncertainty using probability distributions

Roadmap

- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

- What is **examinable**?
  - everything in the videos
  - all "Essential" readings

Roadmap: weeks 2 to 4

- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

- Block 1 **Week 2**
  - Module 1: a brief look at speech production and perception
- Block 1 **Week 3**
  - Foundations: signals
  - Module 2: speech signals and the source-filter model
- Block 1 **Week 4**
  - Foundations: phonetics

Roadmap: weeks 4 to 7

- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

- Block 1 **Week 4**
  - Module 3: text processing
- Block 1 **Week 5**
  - Class trip
  - Module 4: pronunciation & prosody
- Block 1 **Week 6**
  - Assignment Q&A
  - Module 5: waveform generation
- Block 1 **Week 7**
  - Submission of first assignment

Roadmap: weeks 8 to 11

- Modules 1-2: The basics
- Modules 3-5: Speech synthesis
- Modules 6-9: Speech recognition

- Block 2 **Week 8**
  - Foundations: mathematics & probability
  - Module 6: pattern matching
- Block 2 **Week 9**
  - Module 7: feature engineering
  - Module 8: the Hidden Markov Model
- Block 2 **Week 10**
  - Module 9: connected speech
- Block 2 **Week 11**
  - Writing clinic
  - Submission of second assignment

Some of the things we will learn

- Basic digital signal processing
- Text processing - rules *vs* learning from data
- Classification and Regression Trees (CARTs)
- Dynamic programming - Dynamic Time Warping (DTW); the Viterbi algorithm
- Generative models - Hidden Markov Models (HMMs)
- Probability - Bayes' Rule
- Expectation-Maximisation - the Baum-Welch algorithm
- Finite state networks - N-gram language models
- Feature extraction and engineering - Mel Frequency Cepstral Coefficients (MFCCs)

Introduction to the course

- learning outcomes
- delivery
- timetable
- background required
- course coverage
- assignments

Lab facilities

- All practicals are done on computers running Linux
- Appleton Tower 4.02: fully-supported computers, file backup, 24/7 access
  - We'll automatically create accounts for everyone. Look out for announcements.
  - We have ***no*** resources to support use of your own computer (so, for experts only)
- Unix/Linux/command line
  - Basics: terminal, mv, cp, cd, mkdir, starting programs - intro sessions available
- Never switch off machines, just log out
- Lab access
  - Via matriculation card at any time (PIN required out of hours)
  - Follow instructions on Learn to get your card activated