COMP47700
Speech and Audio

Dr Alessandro Ragano

Office: 3rd Floor, Science East, Insight Centre
School of Computer Science
University College Dublin

alessandro.ragano@ucd.ie

1.1 Module Outline

Module Content

1. Introduction to speech and audio processing
2. Basic audio processing
3. Speech
4. The human auditory system
5. Psychoacoustics
6. Speech communications
7. Audio analysis
8. Advanced topics, e.g.
    - Psychoacoustic modelling
    - Speech and audio quality
    - Objective models of intelligibility and quality
    - Speaker recognition
    - Spatial Audio

Learning Outcomes

On successful completion of this module, the learner will be able to

1. Analyse speech and audio signals and features
2. Articulate the characteristics of speech, speech production and speech understanding
3. Describe the signal characteristics of speech and audio signals using appropriate terminology
4. Apply signal processing algorithms to speech and audio signals
5. Create programmes to conduct experiments on speech and audio samples building on third party software libraries

Also... Bigger Picture

Perspectives on Big Data/Predictive Analytics/Machine Learning
Dealing with 'natural' data: capturing real world inputs, cleaning, normalising, classifying...

Unexpected/Bonus learning

1. Python skills
   - wrinkles and configurations
   - powerful libraries for speech and audio processing, math, digital signal processing (DSP), visualisation, data wrangling etc.
2. Human physiology and the mechanics of communication (about yourself)
3. Quality estimation – more aware of compression etc. (you can't unhear it)
4. Subjective testing – biases and the importance of standards when conducting trials (data understanding)

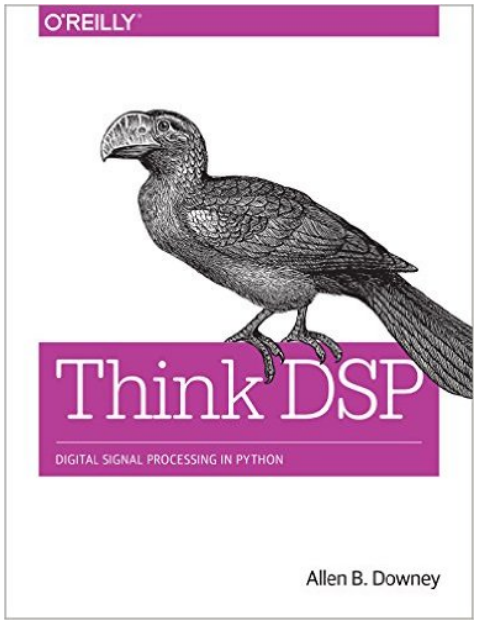
Recommended Textbooks

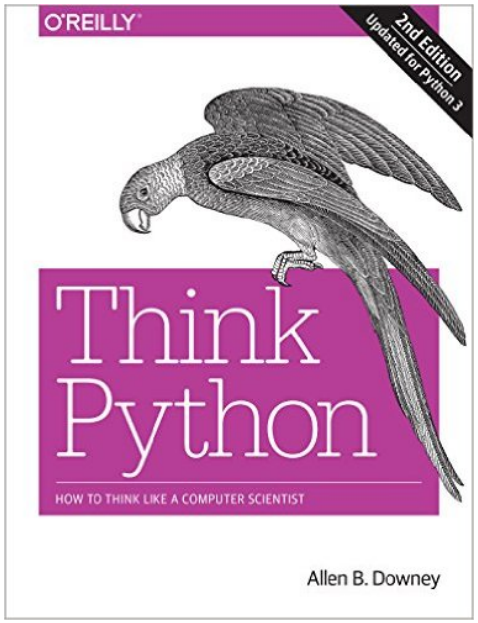
1. **Main Text: Ian Vince McLoughlin (2016), Speech and Audio Processing: A MATLAB-based Approach, 1st edition, Cambridge University Press (PDF available in UCD library)**
2. Ian McLaughlin (2009), Applied Speech and Audio processing, 1st edition, Cambridge University Press
3. Ben Gold, Nelson Morgan, Dan Ellis, (2011), Speech and Audio Signal Processing: Processing and Perception of Speech and Music, 2nd edition, Wiley
4. Lawrence Rabiner, B H Juang (1993), Fundamentals of Speech Recognition, 1st edition, Pearson
5. Lawrence Rabiner, Ronald Schafer (2010), Theory and Applications of Digital Speech Processing, Pearson
6. Ken Pohlmann (2010), Principles of Digital Audio, 6th Edition, McGraw-Hill

Free Online Texts

greenteapress.com

ThinkDSP and ThinkPython by Allen Downey





`https://greenteapress.com/wp/`

How will the module run?

In Person with Brightspace Resources

All Learning Resources: Slides, Colab notebooks, Additional Papers on Brightspace.
All Assignments: Submitted through Brightspace.
All Tutorials/Labs: Thurs 14:00 H1.49-SCH.
Lectures: Wed 10:00 B005-HEA.

Lab: Python for Speech and Audio Processing

Why Python?

Most common languages in the audio industry:

- C++ - real-time audio processing e.g., headphones with active noise cancellation
- Python - Audio research, machine/deep learning for audio, foundation models APIs e.g., speech recognition, text-to-speech, etc

Using Python and its machine learning capabilities, we can build practical and complete applications for a wide range of speech and audio processing tasks.

Python Libraries

Numpy, SciPy, matplotlib, Jupyter, anaconda, librosa, etc.

- Numpy – Numerical array handling for python (i.e. taking the for loops out of matrix algebra computation)
- SciPy – SciPy is a collection of mathematical algorithms and convenience functions built on the Numpy extension of Python
- matplotlib – visualisation and scientific graphics for python
- Jupyter – IPython interactive shell and Jupyter Notebooks (interactive python coding, visualisation, and media)
- anaconda[1] – everything python for DSP and data scientists (Python, Jupyter Notebook, numpy, scipy, and matplotlib in one)
- librosa – audio and speech analysis toolbox [2]

[1] `https://www.anaconda.com/distribution/`
[2] `http://librosa.github.io`

Lab 1

This aim of this lab is to familiarise yourself with:

- Colab: Google based Jupyter notebooks and Librosa Library
- Functionality available in libraries and where to find the documentation
- Different syntax and ways of doing the same thing
- Basic mathematical representations in python with numpy
- Simple audio processing – reading and writing files
- Visualisation – basic plots with matplotlib

Lab 2: Basic audio processing in Python

Reading a speech recording from a WAV file
Playing sounds
Representing a signal digitally
Plotting the signal – e.g. visualisation of time series, spectrograms

Assessment

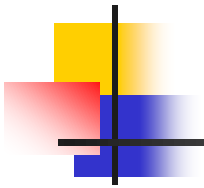
Assignments

100% of module mark is CA. Labs, Quizzes, and a Term Project.

Assignment		
6 Labs Notebook	Individual	40%
Term Project	Pairs/Individual	40%
Midterm In-Class Quiz	Individual	10%
Term Online Quiz	Individual	10%

Supplemental Info/Interviews

If there is a suspicion of plagiarism, use of ChatGPT/LLMs or assignment that are not the work of a student, 0/NM awarded, request for supplemental submission and/or interviews may be used to supplement grading assessment.



Plagiarism & UCD Computer Science

- **Plagiarism is the "failure to cite or otherwise acknowledge ideas or phrases used in any paper, exercise, assessment or project submitted in a course but gained from another source, such as a published text, another person's work, or materials on the internet" [UCD Academic Integrity Policy]**
  - **An example of academic misconduct: "any attempt by someone to seek unfair advantage in relation to [an] academic activity or which facilitates others to gain an unfair advantage, or to profit from the sharing or selling of your own or others' work without permission"**

Student Conduct and Academic Integrity **School of CS Plagiarism Policy**

- CS staff/demonstrators are **proactive** in looking for plagiarism → **CS Plagiarism committee**
- Student who *enables* plagiarism is *equally responsible* for it
- **Examples** of plagiarism:
  - Copying some/all of the work of another student and submitting it as your own work
  - Copying some/all of an assignment from the Internet/book/etc without referencing it
  - Sharing individual work with another student (by e-mail, FB messenger, WhatsApp, …)
  - Making your work available (on GitHub, website, social media, …) before lecturer gives permission
  - Submission of AI-generated content without explicit permission and attribution
  - Students collaborating at too detailed a level e.g. consulting each other after implementing a line/block/segment of code and sharing the results, then individually submitting the resulting work

UCD Dignity and Respect Support Service

Supporting any student, employee or UCD community member who is impacted by issues of **bullying, harassment or sexual misconduct**

- **Confidential and proactive support**
- **Specialist information and guidance**
- **Support to explore informal and formal resolution options**
- **Ongoing aftercare support**
- **Drop in or make an appointment**

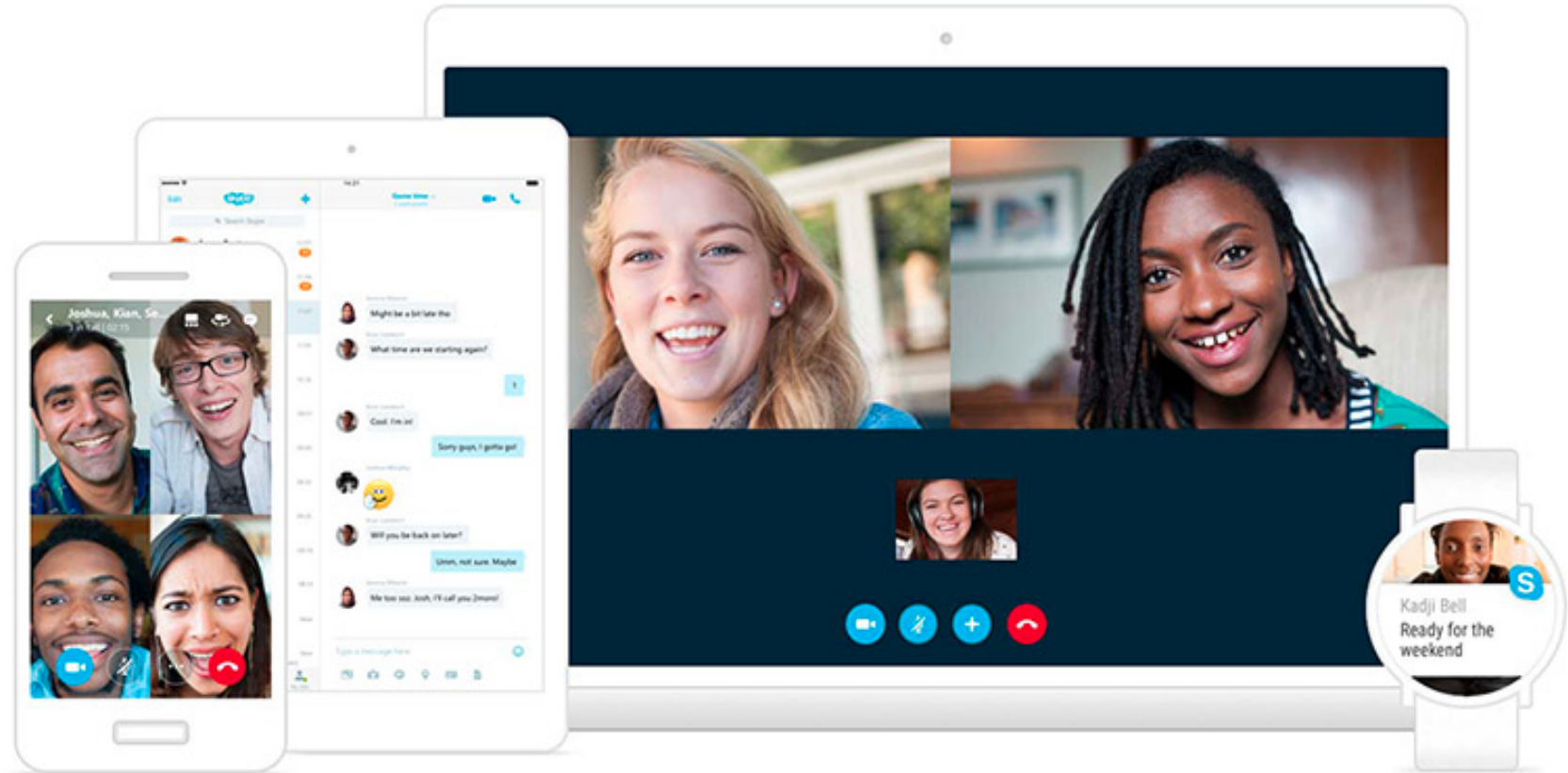
L532, Level 5 James Joyce Library Building

respect@ucd.ie **ucd.ie/dignityandrespect**

Report + Support Tool

1.2 Speech and Audio Introduction

Computers and Audio: VoIP



Computers and Audio: Streaming Media

Computers and Audio: Voice Assistants

Computers and Audio: Hearing Aids

Speech Communications: We've come a long way...

Source: https://en.wikipedia.org/wiki/Old_Stockholm_telephone_tower

Summary: Speech and Audio Processing

Context

Framework and Focus of the module: Theory and Practical

Module Framework

Fundamental Theories

Acoustics, Linguistics, Digital Signal Processing (DSP), Speech and Audio Production and Perception

Signal Representation

Temporal, Spectral, homomorphic, LPC

Algorithms

Speech/Silence detection (VAD), Voiced/Unvoiced, Pitch/Formant

Applications

ASR, Speech and Audio Coding, Synthesis, Verification (Speech/Speaker/Audio), Language Translation

Module Focus: Concepts and Practice

Worth saying again...

This is a module for *computer scientists*
It is NOT a module in digital signal processing (DSP). It is NOT a module in automatic speech recognition (ASR).

Theory

Less: Maths, Derivations, DSP Fundamentals

Concepts

More: How to apply theory to the domain

Practice and Apply Theory Offline

Using Python to implement worked examples of the concepts based on theory.

In Four Slides

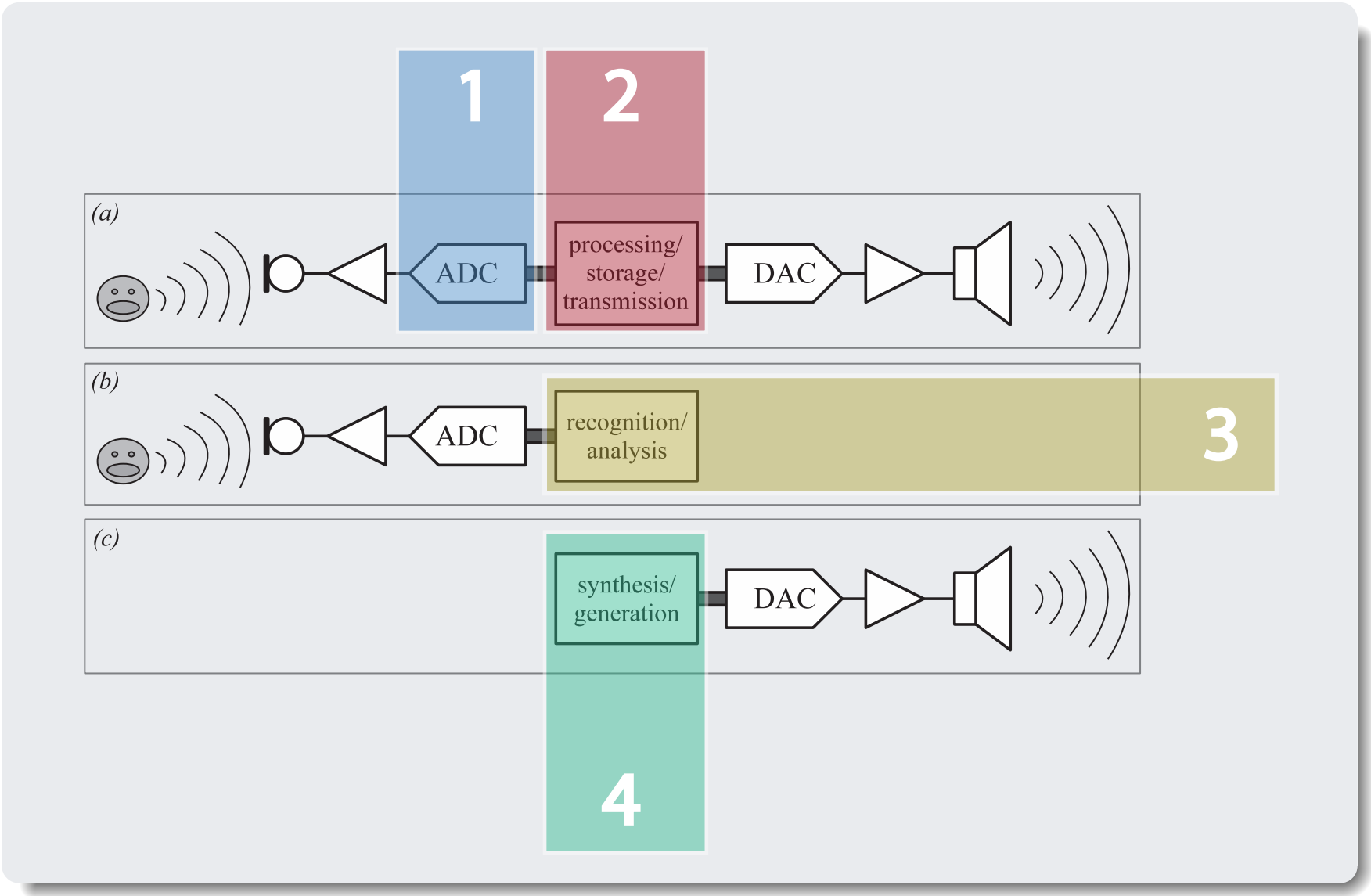
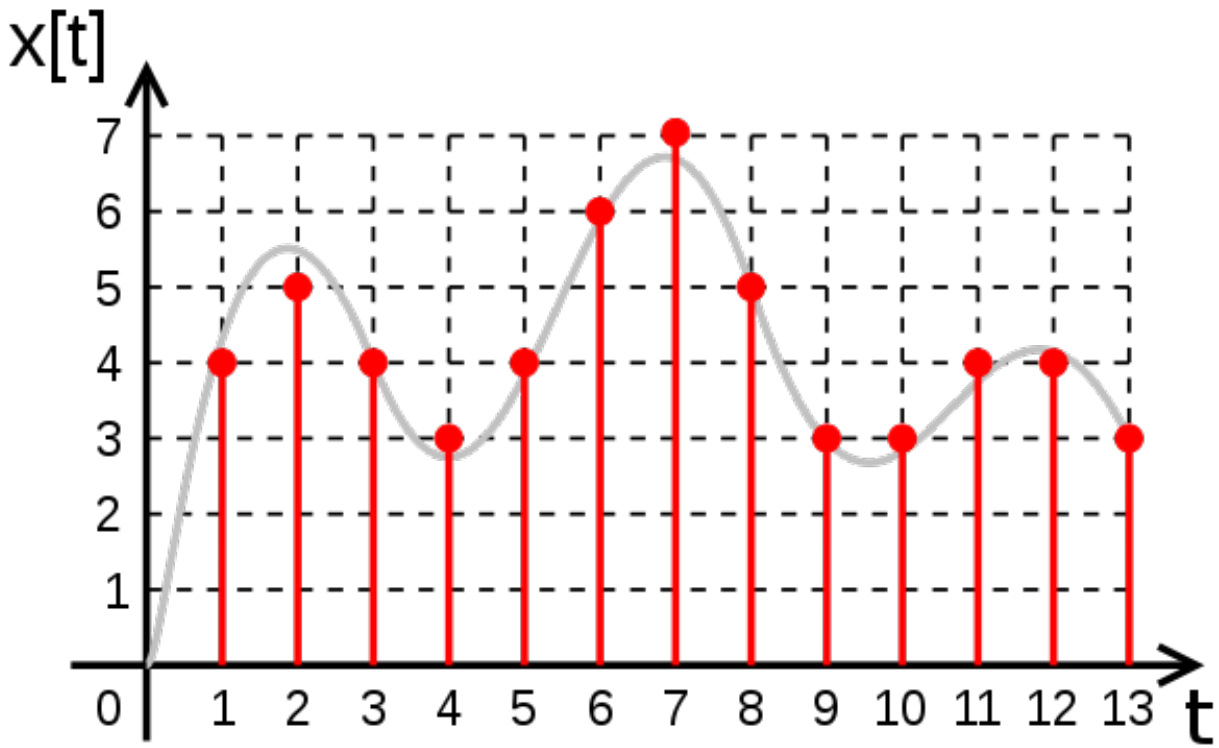


Figure adapted from McLaughlin (2009, p.3)

1: Digital Representation of Sound



2: Processing

Speech Effects

e.g. noise removal, speech codecs e.g. mp3, pitch shifting

Audio Effects

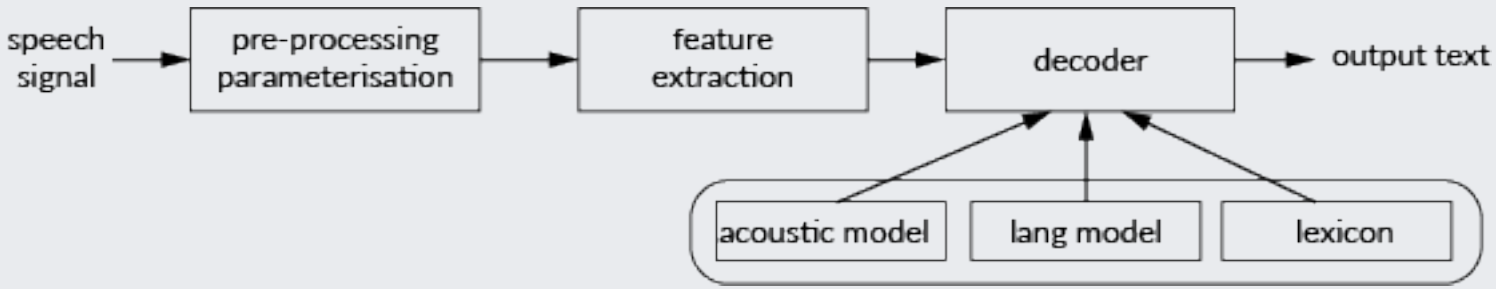
e.g. reverb, mixing (soundtrack + dialogue movie), virtual spatial audio

Understanding the theory behind audio perception and hearing is important for this

3: Analysis (and Recognition)

Quality vs. Intelligibility, Speaker Identification, Emotion Recognition

Automatic Speech Recognition (ASR) or speech-to-text

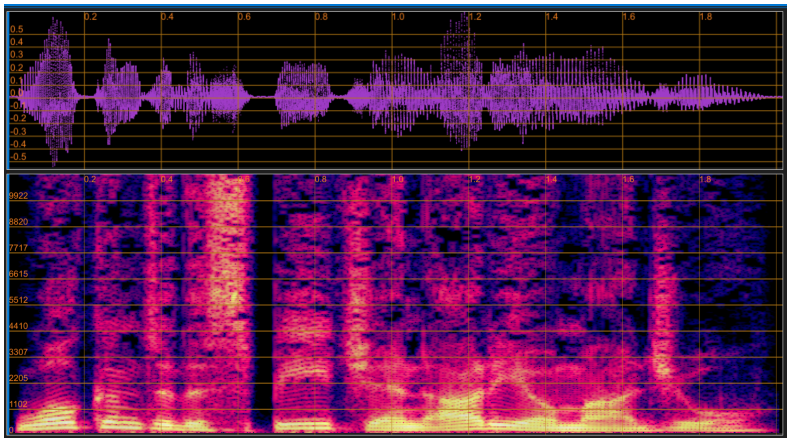


Understanding the theory behind speech production, language and phonetics is important for this

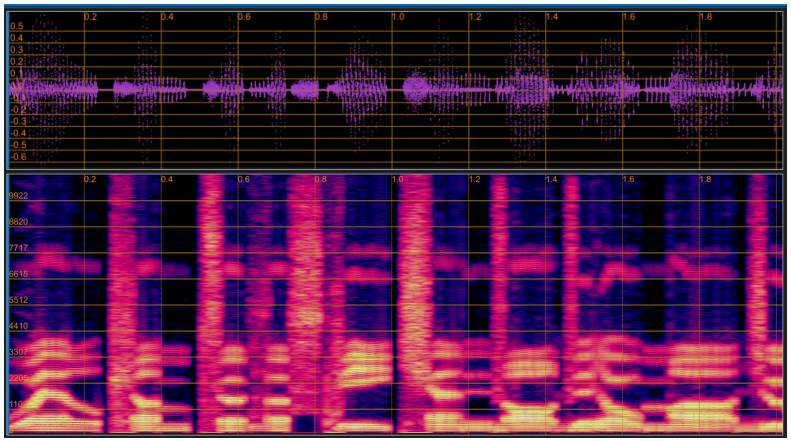
4: Synthesis

Text to Speech, Customer Support Lines

e.g. Stephen Hawking, navigation systems, automatic menus



Speech 1



Speech 2

Welcome to the speech and audio module