ENCS539
Spoken Language Processing

Second Semester – 2021

Abualsoud Hanani

Objectives

- To understand basic principles of human speech production and perception
- To apply basic techniques for visualisation and analysis of speech signals
- To understand basic, practical phonetics
- To demonstrate in-depth knowledge of most common feature extraction techniques, e.g. MFCC, LPCC, etc.
- To apply the most popular modelling techniques in speech processing applications such as VQ, GMM, HMM, Deep Neural Networks, etc.

Course objectives …continue

- To understand in-depth automatic speech recognition using hidden Markov models

- To understand practical issues in automatic speech processing and integration of multiple modalities
- To understand the basic principals of speech synthesis
- To compare between the most common speech synthesizers
- To understand the principals of speech coding

Assessment

- 1. Participation and attendance …… 5%
- 2. Assignment ……..10%
- 3. Term project ……20%
- 4. Mid-term Exam ……25%
- 5. Final exam ……40%

Books

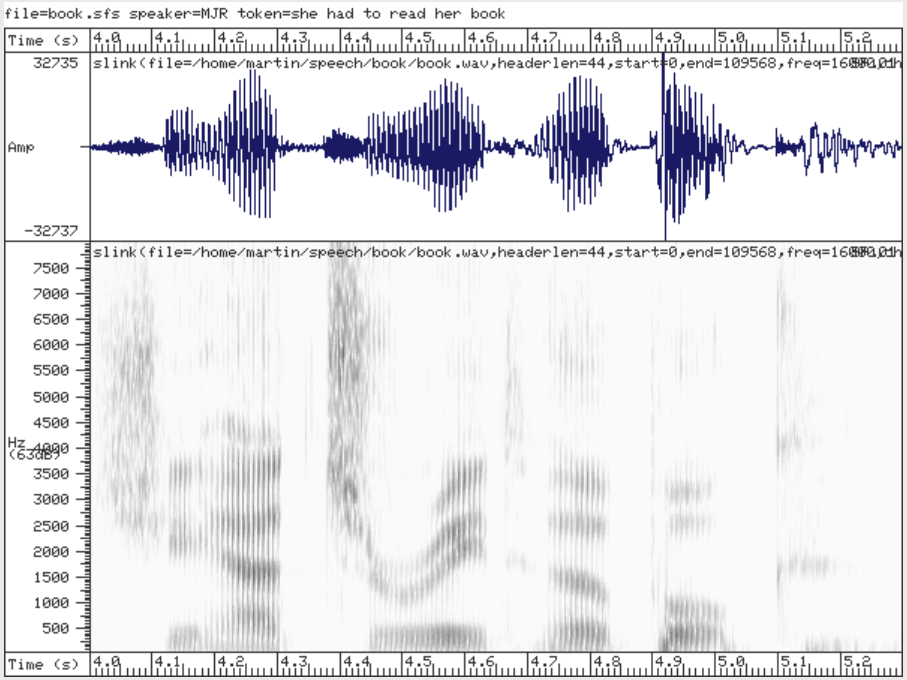
- J N Holmes and W J Holmes, *"Speech Synthesis and Recognition"* Second Edition, 2001
- L Rabiner and B-H Juang, "*Fundamentals of Speech Recognition*", Prentice Hall, 1993
- L R Rabiner and R W Schafer, "*Digital Processing of Speech Signals*", Prentice-Hall, 1978
- B Gold and N Morgan, *"Speech and Audio Signal Processing",* Wiley, 2000
- J.R. Deller, J.H.L. Hansen, J.G. Proakis, "*Discrete-Time Processing of Speech Signals*", IEEE Press, 1999

More Books

- G T Altmann, "*The Ascent of Babel*", Oxford University Press, 1998
- K Jelinek, "*Statistical Methods for Speech Recognition*", MIT Press, 1998
- Steve Young, "*The HTK Book*", Cambridge University Engineering Department

Software used during course

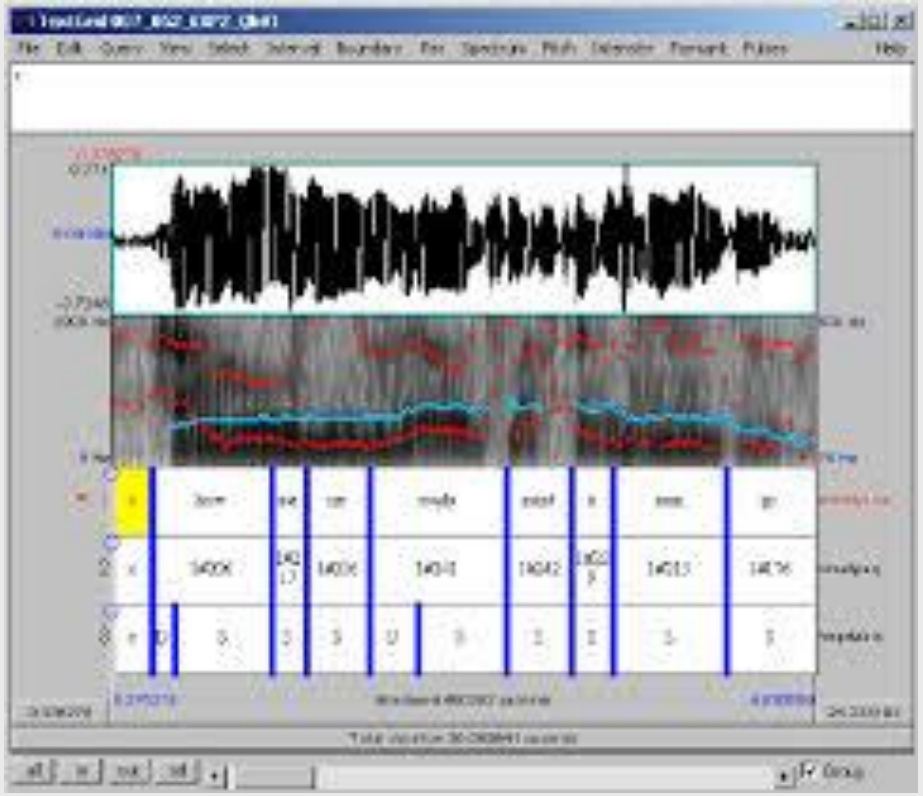
- SFS – Speech Filing System, by Mark Huckvale, available **free** from UCL website http://www.phon.ucl.ac.uk/resource/sfs/



Praat Toolkit

- **Praat**: doing phonetics by computer

  http://www.fon.hum.uva.nl/praat/



Kaldi toolkit

- Toolkit for speech recognition written in C++
- Includes implementation of most kind of deep neural networks
- Use GPU implementation
- Command line interface
- Python friendly

HTK - toolkit

- HTK – “Hidden Markov Model Tool Kit”
- ‘Toolkit’ to build conventional HMM-based speech recognition systems – used in labs worldwide
- Applicable to other speech technology applications
- Off-line evaluation
- Available free from Cambridge University website http://htk.eng.cam.ac.uk/

Matlab and Python

Matlab toolbox's:

- Signal processing toolbox
- Voicebox toolbox
- Netlab toolbox

- WEKA- an open source modelling toolkit

Speech Science

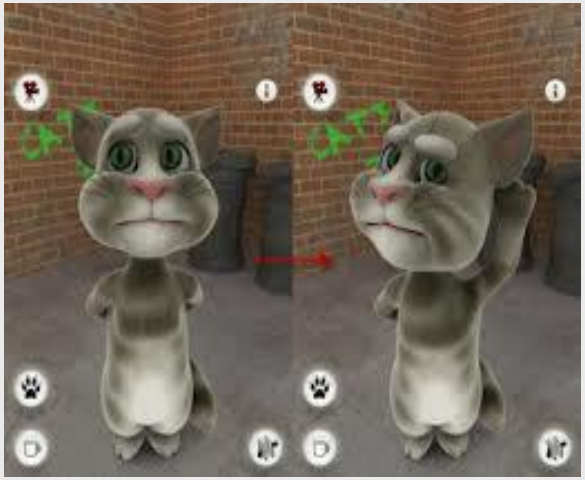
- What is speech?
  - Unique
  - Human-to-human communication
- Multi-disciplinary
  - Acoustics
  - Phonetics and linguistics
  - Psychology
  - Biology and neuroscience
  - Computer science, cognitive science, AI
  - Mathematics and statistics
  - Electronic engineering

Spoken Language Technology

- Speech technology is concerned with:
  - Automatic analysis of human speech
  - Human interaction with computers and other technology, using spoken language
  - Integration of speech with other modalities
- It has potential applications wherever people interact with machines or use spoken language

Application areas

- Education
- Entertainment
- Forensics & crime prevention
- Information retrieval
- Biometrics
- Help to disabled people



Advantages of Speech

- Fast
- Natural
- Rich and Complex
- Works over the telephone
- “Hands and Eyes Free”
- Works in hostile environments
- Needs minimal panel space

Component Technologies

- Automatic speech recognition (speech-to-text)
- Speech synthesis (text-to-speech)
- Speech coding (parameters/samples represent speech)
- Spoken language understanding (content meaning)
- Spoken dialogue processing

Component Technologies (contd)

- Paralinguistic speech processing
  - Speaker verification / recognition / identification (who is speaking?)
  - Language recognition (what language?)
  - Gender recognition (male/fmale)
  - Accent/Dialect recognition (which accent/dialect?)
  - Age estimation from speech
  - Topic spotting
- Speech verification

Summary

- Information on the course
- Introduction to spoken language processing
- and speech technologies
  - Advantages
  - Applications areas
  - Multi-disciplinarity