

PUE 4113 Speech Processing

Lecturer: Prof. Ciira Maina
E-mail: ciira.maina@dkut.ac.ke
Tel: +254(0)716196331

Summary

The speech signal contains significant information. In addition to information about what was said and who said it, we can determine the emotional state of the speaker, and if recording a conversation, we can determine the level of engagement of each of the participants.

Conversational interfaces such as Alexa and Siri allow users to interact with complex systems through voice. These systems show that speech technology is now mature and capable of deployment in real world systems.

This course will introduce the student to the fundamentals of speech processing with a focus on the theory and deployment of real world systems.

Course content

1. Speech production and perception
2. Speech signal analysis
3. Feature extraction
4. Modeling speech
5. Speech coding
6. Speech systems: Speech recognition, speaker recognition
7. Machine learning for speech processing
 - Gaussian Mixture Models
 - Hidden Markov Models
 - Neural Networks - Deep neural networks, generative models, sequence-to-sequence models

Prerequisites

In this course we will build speech systems using Python.

To install python see <https://www.anaconda.com/distribution/>.

Also, for an introduction to python see <https://github.com/ciiram/ml-intro-notes/blob/master/notebooks/Introduction-to-Python.ipynb>

As such you need to have a Python system installed with the following tools.

- librosa <https://librosa.github.io/librosa/>
- sklearn <https://scikit-learn.org/stable/>
- Tensorflow <https://www.tensorflow.org/>
- Keras <https://keras.io/>

In addition, the software package Audacity <https://www.audacityteam.org/> will be useful to obtain recordings.

Course books

I will use the following books as references:

- Rabiner, L. R., & Schafer, R. W. (2007). Introduction to digital speech processing. Foundations and Trends® in Signal Processing, 1(1-2), 1-194.
- Huang, X., Acero, A., Hon, H. W., & Reddy, R. (2001). Spoken language processing: A guide to theory, algorithm, and system development (Vol. 1). Upper Saddle River: Prentice hall PTR.
- Bishop, C. M. (2006). Pattern recognition and machine learning. springer.

Assessment

There will be continuous assessment tests during the semester and a final exam. The dates will be announced at a later date.