BIRZEIT UNIVERSITY

Faculty of Engineering and Technology

Electrical and Computer Engineering

Spoken Language Processing (ENCS5344)

<u>Second Semester 2024</u>

Instructor: Dr. Abualsoud Hanani

Office: Masri222 Office Hours: *See Ritaj*

- **pre-requisite**(s): Fundamental knowledge of propability theory and digital signal processing.

- **Course describtion**

This course builds directly on students’ skills and knowledge in signal processing gained during ENEE2302 and ENCS4310 courses. It aims to reinforce concepts learned in those courses, to introduce new tools needed to deal with time-varying signals and to have students apply what they have learned to their own voices. Speech processing methodologies will be covered in lectures, assignments and course project. The course's major target is to provide students with the knowledge of basic characteristics of speech signal in relation to production and hearing of speech by humans, describe basic algorithms of speech analysis common to many applications, give an overview of applications (recognition) and to inform about practical aspects of speech algorithms implementation.
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 entirel digital, hence only discrete-time theory is presented in this course.

❖ **Course Aims and objectives**:

This course aims to:

a) Familiarise students with modeling the vocal tract as a digital, linear time-invariant system.
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 students with the practical aspects of speech processing, including robustness, and applications of speech processing, including speech enhancement, speaker recognition and speech recognition.
e) Give students practical experience with the implementation of several components of speech processing systems.

❖ **Intended Learning Outcomes (ILOs')**:

Upon successful completion of this course, students should be able to:

1. Express the speech signal in terms of its time domain and frequency domain representations and the different ways in which it can be modeled.

2. Derive expressions for simple features used in speech classification applications.

3. Explain the operation of example algorithms covered in lectures, and discuss the effects of varying parameter values within these.

4. Synthesize block diagrams for speech applications, explain the purpose of the various blocks, and describe in detail algorithms that could be used to implement them.

5. Implement components of speech processing systems, including speech recognition and speaker recognition, in MATLAB/Python.

6. Deduce the behavior of previously unseen speech processing systems and hypothesise about their merits.

❖ Assessment Criteria:

2. HWs & Projects ……..………………..20%

4. MidTerm Exam ………………………..35%

5. Final exam ……………………………..45%

❖ Course Resources:

[1] Martin Russell, "Spoken Language Processing lecture notes", EECE, University of Birmingham, UK, version 11, 2008.

[2] Rabiner, L. R., and Juang, B.H. ,"Fundamentals of Speech Recognition", Prentice-Hall, New Jersey, 2009.

[3] John Holmes and Wendy Holmes, "Speech Synthesis and Recognition, 2nd Edition", Taylor & Francis, 2001

[4] Deller, John R., John G. Proakis, and John HL Hansen. Discrete-time processing of speech signals. New York: Macmillan publishing company, 1993.

❖ Course Schedule:

Week No	Topic	Reading
1	Introduction to the course and the outline	
2	Introduction to Speech processing technologies and apps	
3	Fundamentals of Speech Science: Speech Production Mechanism, Sound Units, Acoustic Theory, Digital Modeling	
4	Time domain analysis of speech signal: Short-time analysis, frame, short time energy, short time zero-crossing, short time average magnitude, short time autocorrelation, silence removing, pitch detection, voice/voiceless classification.	
5	Frequency domain analysis of speech signal: short Fourier transform, Short time Spectrum	
6	Linear Prediction Analysis: what is it good for?, Prediction of a sample from past samples, linear prediction (LP), Error of LP, Determination of vocal tract characteristics using LP analysis, Spectrum estimated by LP. Features derived from LP.	
7	Frontend Processing (MFCC and LPCC)	
	Mid-Term Exam	
8	Vector Quantization and K-means	
9	Gaussian Mixture Model (GMM)	
10	Speaker, language and Dialect Recognition	
11	Introduction to Hidden Markov Model (HMM)	
12-13	Automatic Speech Recognition (ASR)	
14	Introduction to Speech Processing with Deep Learning	
15	Project discussion and presentation	
16	Final Exam	