Subject Description Form

Subject Code	EIE558
Subject Title	Speech Processing and Recognition
Credit Value	3
Level	5
Pre-requisite/ Co-requisite/ Exclusion	Nil
Objectives	This subject aims to enable students to master the state-of-the-art theories and technologies behind various speech-related products and services, such as mobile phones, voice search, Internet phones, dialog systems, voice biometrics, and voice cloning. The course will cover theoretical foundations, algorithms, and practical issues of speech processing and recognition systems. The course emphasizes how recent advances in deep learning and deep neural networks revolutionize these systems. After completing the subject, students will understand what the current speech technologies can offer and be able to apply speech processing techniques to industrial and commercial applications. The course is suitable for students with a background in signal processing and statistics. It is also ideal for research students working in speech processing. Prior experience in speech processing is not necessary.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: a. master the fundamental principles behind voice-enable products and services; b. know what the current state-of-the-art speech technologies can offer; c. apply speech processing technologies to voice-enabled products and services; d. take the limitations of current speech technologies into consideration when deploying voice-enabled services.
Subject Synopsis/ Indicative Syllabus	**Part I: Fundamental Concepts** 1. Speech Production and Modelling 1.1 Physiology of speech generation; acoustic characteristics of speech sounds 1.2 Discrete-time speech production model 2. Speech Analysis and Parameterization 2.1 Short-time Fourier analysis; spectrograms 2.2 Linear prediction; cepstrum; LPCC; MFCC **Part II: Advanced Topics and Applications** 3. Speech Enhancement 3.1 Spectral subtraction; 3.2 DNN-based approaches 4. Speech Coding 4.1 Attributes of speech coders and coding standards 4.2 Waveform coding: PCM and ADPCM 4.3 Linear predictive coding: LPC and MELP 4.4 Analysis-by-synthesis coders: CELP and MPLPC 5. Machine Learning and Deep Learning 5.1 Gaussian mixture models 5.2 Support vector machines 5.3 Deep Learning and deep neural networks 5.4 Convolutional neural networks, ResNet, and DenseNet 6. Speech Recognition 6.1 Types of speech recognition 6.2 Hidden Markov models (HMM); language models; DNN-HMM 6.3 End-to-End speech recognition: Seq2Seq and CTC 6.4 Speaker adaptation: MAP; MLLR; DNN adaptation 7. Speaker Recognition

7.1 Types of speaker recognition
7.2 Speaker modelling: GMM-UBM and GMM-SVM
7.3 Speaker embedding: i-vectors; x-vectors; ResNet and DenseNet speaker embeddings
7.4 Scoring: LDA, PLDA, and cosine distance
7.5 Performance metrics: EER, minimum DCF and actual DCF

Teaching/Learning Methodology

The theories and applications of various speech technologies will be discussed and explained in lectures. Lab sessions will be provided to strengthen students' understanding on the theories and hands-on experiences. Students will also be requested to write an essay of a selected topic.

Teaching/Learning Methodology	Intended Subject Learning Outcomes			
	a	b	c	d
Lecture	✓	✓	✓	✓
Tutorial	✓			
Laboratory			✓	✓
Essay writing	✓	✓		

Assessment Methods in Alignment with Intended Learning Outcomes

Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)			
		a	b	c	d
1. Laboratory reports	30%	✓		✓	
2. Quiz	10%	✓			
3. Essays	20%		✓		✓
4. Examination	40%	✓	✓		✓
Total	100%				

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:

1. **Lab Reports**: For each lab session, students will need to understand the fundamental concepts [Outcome (a)] before they can complete the lab exercises and write a report. Because the lab sessions involve the application of speech technologies [Outcome (c)], students' ability to apply these technologies should be reflected in their reports.
2. **Quiz:** A quiz will be given to check students' understanding on the fundamental concepts.
3. **Essays:** Students will need to conduct surveys on various speech technologies, find out the limitations of these technologies [Outcome (d)], and determine what the current technologies can offer [Outcome (b)].
4. **Exam:** Students will need to answer questions about the fundamental concepts [Outcome (a)] of various speech technologies and their applications [Outcome (b)]. Limitations of current speech technologies [Outcome (d)] will also be asked in the exam.

Student Study Effort Expected

Class contact:	
▪ Lectures and tutorials	30 Hrs.
▪ Laboratory sessions	9 Hrs.
Other student study effort:	
▪ Writing essay	22 Hrs.
▪ Writing laboratory report and self learning	45 Hrs.
Total student study effort	106 Hrs.

Reading List and References

1. M.W. Mak and J.T. Chien, "*Machine Learning for Speaker Recognition*", Cambridge University Press, 2020.
2. Z. Bai and X.L. Zhang, "Speaker recognition based on deep learning: An overview," *Neural Networks*, vol. 140, pp. 65-99, 2021.

3. T. Backstrom, *Speech Coding: With Code-Excited Linear Prediction*, Springer, 2017.
4. S. Watanabe and J.T. Chien, "*Bayesian Speech and Language Processing*", Cambridge University Press, 2015.
5. J. Benesty, et al. *Speech Enhancement*, Academic Press, 2014.
6. Y. LeCun, Y. Bengio and G.E. Hinton, "*Deep Learning*", Nature, vol. 521, pp. 436-444, May 2015.
7. T. Kinnunen and H. Z. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication*, 2010.
8. J.R. Deller, J.G. Proakis, and J.H.L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Pub. Company, 2000.
9. L.R. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
10. S.Y. Kung, M.W. Mak and S.H. Lin, *Biometic Authentication: A Machine Learning Approac*h, Prentice Hall, 2005.
11. A.M. Kondoz, *Digital Speech: Coding for Low Bit Rate Communications Systems*, 2nd Edition, Wiley, 2004.
12. T.E. Quatieri, *Discrete-Time Speech Signal Processing*, Prentice Hall, 2002.

July 2022