<table>
<thead>
<tr>
<th>Subject Code</th>
<th>EIE558</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Title</td>
<td>Speech Processing and Recognition</td>
</tr>
<tr>
<td>Credit Value</td>
<td>3</td>
</tr>
<tr>
<td>Level</td>
<td>5</td>
</tr>
<tr>
<td>Pre-requisite/ Co-requisite/ Exclusion</td>
<td>Nil</td>
</tr>
</tbody>
</table>

**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, voice over IP, Internet phones, directory services, and voice biometrics. The course will cover theoretical foundations, major algorithms and practical issues of state-of-the-art speech processing and recognition systems. After completing the subject, students will have the understanding of 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 background in signal processing and statistics. It is also suitable 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; phonemics and phonetics; acoustic characteristic of speech sounds
   1.2 Discrete-time speech production model

2. Speech Analysis and Feature Extraction
   2.1 Short-term processing: short-term Fourier analysis; spectrograms
   2.2 Speech parameterization: linear prediction; cepstrum; LPCC; MFCC

**Part II: Advanced Topics and Applications**

3. Speech Enhancement
   3.1 Classical: Spectral subtraction; Wiener filtering; MMSE estimator
   3.2 Advanced: DNN-based approaches

4. Speech Coding
   4.1 Attributes of speech coders; waveform coding; LPC; MELP
   4.2 Analysis-by-synthesis coders: CELP, MPLPC, G.729, G.723.1

5. Speech Recognition
   5.1 Feature parameters for speech recognition
   5.2 Continuous speech recognition: hidden Markov models (HMM), language models, deep neural networks (DNN), DNN-HMM hybrid
   5.3 Speaker adaptation: MAP, MLLR, speaker adaptive training

6. Speaker Recognition
   6.1 Applications of speaker recognition technologies
   6.2 Classical methods: Feature warping, GMM-UBM
   6.3 Advanced methods: GMM-SVM, i-vectors, PLDA, and DNN i-vectors
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.

<table>
<thead>
<tr>
<th>Teaching/Learning Methodology</th>
<th>Intended Subject Learning Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture</td>
<td>a</td>
</tr>
<tr>
<td>Tutorial</td>
<td>✓</td>
</tr>
<tr>
<td>Laboratory</td>
<td></td>
</tr>
<tr>
<td>Essay writing</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Assessment Methods in Alignment with Intended Learning Outcomes**

<table>
<thead>
<tr>
<th>Specific assessment methods/tasks</th>
<th>% weighting</th>
<th>Intended subject learning outcomes to be assessed (Please tick as appropriate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Laboratory reports</td>
<td>30%</td>
<td>✓</td>
</tr>
<tr>
<td>2. Essays</td>
<td>25%</td>
<td>✓</td>
</tr>
<tr>
<td>3. Examination</td>
<td>45%</td>
<td>✓</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

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. **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 (c)].

3. **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 20 Hrs.
- Writing laboratory report and self learning 43 Hrs.

Total student study effort 102 Hrs.

**Reading List and References**
