

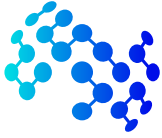
Core Courses Syllabi

NLP706 - Advanced Speech Processing

Title	Advanced Speech Processing
Code	NLP706
Loading	4 Credit-hours
Prerequisites	<ul style="list-style-type: none"> • NLP 703 Speech Processing • Understanding of calculus, algebra, probability and statistics • Programming in Python or similar language
Catalog Description	This course focuses on developing skills for performing research to advance the state of the art in Speech Processing. This course builds upon concepts from Basic Speech Processing (NLP 703) and assumes familiarity with fundamental concepts in Speech Recognition, Speech Synthesis and Speaker Identification.
Goal	This graduate course aims to inculcate a deeper understanding of the advanced Speech Processing methods, so the students are capable of researching, developing, and implementing these methods for solving real-world problems. Additionally, a significant goal of this course is to enhance students' teamwork skills by requiring them to participate in group projects.
Content	The course covers three modules: (I) Speech Recognition, (II) Speech Synthesis, (III) Speaker Identification
Recommended Textbooks	<ol style="list-style-type: none"> 1. Chris Manning et al, <i>Foundation of statistical natural language processing</i>, MIT Press, 1999. 2. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. <i>Deep Learning</i>, MIT Press, 2016.
Recommended References & Supplemental Material	<p>Relevant research papers, tech reports, and surveys for each topic, where needed, are identified in the teaching plan ahead. In addition, the following textbooks may be useful:</p> <p>C. Bishop, <i>Pattern Recognition and Machine Learning</i>, Berlin: Springer-Verlag, 2006.</p>



Teaching Week	Topics
1	<p>Speech Recognition</p> <p>Lectures</p> <ul style="list-style-type: none">• Overview of Speech Recognition Frameworks• Relevant papers and assigned reading:<ul style="list-style-type: none">- Ravanelli, et al., <i>The PyTorch-Kaldi Speech Recognition Toolkit</i>. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019- Pratap, Vineel, et al. <i>wav2letter++: The Fastest Open-source Speech Recognition System</i>, 2018- Amodei, Dario, et al. <i>Deep speech 2: End-to-end speech recognition in English and Mandarin.</i>, International conference on machine learning., 2016- Kuchaiev, Oleksii, et al., <i>Mixed-precision training for NLP and speech recognition with openseq2seq.</i>, 2018 <p>Lab</p> <ul style="list-style-type: none">• Discussion on choosing a relevant paper to implement for the project• Start project-1 work
2	<p>Speech Recognition</p> <p>Lecture</p> <ul style="list-style-type: none">• Speech Recognition – Multi-Speaker Speech Recognition• Discussion of papers:<ul style="list-style-type: none">- Dong Wang et al., <i>Deep Speaker Verification: Do we need End to End?</i>, 2017- Quan Wang et al., <i>Speaker Diarization with LSTM</i>, 2017- Amirsina Torfi et al., <i>Text-Independent Speaker Verification Using 3D Convolutional Neural Networks</i>, 2017 <p>Lab</p> <ul style="list-style-type: none">• Continue project-1 work
3	<p>Speech Recognition</p> <p>Lecture</p> <ul style="list-style-type: none">• Speech Recognition – Multilingual Speech Recognition• Discussion of papers:<ul style="list-style-type: none">- Shubham Toshniwal et al., <i>Multilingual Speech Recognition with a Single End-To-End Model</i>, 2018- Dario Amodei et al., <i>Deep Speech 2: End-to-End Speech Recognition in English and Mandarin</i>, 2015 <p>Lab</p> <ul style="list-style-type: none">• Continue project-1 work



Teaching Week	Topics
4	<p>Speech Recognition</p> <p>Lecture</p> <ul style="list-style-type: none">• Speech Recognition – Domain-specific Speech Recognition• Discussion of papers:<ul style="list-style-type: none">- P Bell et al., <i>Complementary Tasks for Context-Dependent Deep Neural Network Acoustic Models</i>, 2015- P Bell et al., <i>Multi-Level Adaptive Networks in Tandem and Hybrid ASR Systems</i>, 2013 <p>Lab</p> <ul style="list-style-type: none">• Preparation of presentation on project-1 work• Continue project-1 work
5	<p>Speech Recognition</p> <p>Assessment 1.1</p> <ul style="list-style-type: none">• Presentation of the projects by different groups <p>Lab</p> <ul style="list-style-type: none">• Peer review of project reports <p>Assessment 1.2</p> <ul style="list-style-type: none">• In-class exam covering module I – Speech Recognition
6	<p>Speech Synthesis</p> <p>Lecture</p> <ul style="list-style-type: none">• Overview in Speech Synthesis Frameworks• Discussion of papers:<ul style="list-style-type: none">- Ping, Wei, et al. <i>Deep voice 3: Scaling text-to-speech with convolutional sequence learning.</i>, 2018.- Shen, Jonathan, et al. <i>Natural TTS synthesis by conditioning Wavenet on MEL spectrogram predictions</i>. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.- Zhizheng Wu et al., <i>Merlin: An Open Source Neural Network Speech Synthesis System</i>, in Proc. 9th ISCA Speech Synthesis Workshop, 2016. <p>Lab</p> <ul style="list-style-type: none">• Discussion on choosing a relevant paper to implement for the project• Start Project-2 work



Teaching Week	Topics
7	<p>Speech Synthesis</p> <p>Lecture</p> <ul style="list-style-type: none">• Speech Synthesis – Multi-speakers Speech Synthesis• Discussion of the papers:<ul style="list-style-type: none">- Ye Jia et al., <i>Transfer Learning from Speaker Verification to Multispeaker Text-to-Speech Synthesis</i>, 2019 <p>Lab</p> <ul style="list-style-type: none">• Continue Project-2 work
8	<p>Speech Synthesis</p> <p>Lecture</p> <ul style="list-style-type: none">• Speech Synthesis – Multilingual Speech Synthesis• Discussion on papers:<ul style="list-style-type: none">- Bo Li et al., <i>Bytes are All You Need: End-to-End Multilingual Speech Recognition and Synthesis with Bytes</i>, 2018- Alexander Gutkin et al., <i>Areal and Phylogenetic Features for Multilingual Speech Synthesis</i>, 2017 <p>Lab</p> <ul style="list-style-type: none">• Continue Project-2 work
9	<p>Speech Synthesis</p> <p>Lecture</p> <ul style="list-style-type: none">• Speech Synthesis – Domain-specific Speech Synthesis• Discussion of papers:<ul style="list-style-type: none">• Atharva Chogle et al., <i>Text to Speech Conversion for Vocally Disable</i>, 2017• Hay Mar Htun et al., <i>Text to Speech Conversion Using Different Speech Synthesis</i>, 2015 <p>Lab</p> <ul style="list-style-type: none">• Preparation of presentation on project-2 work• Continue project-2 work
10	<p>Speech Synthesis</p> <p>Assessment 2.1</p> <ul style="list-style-type: none">• Presentation of the projects by different groups <p>Lab</p> <ul style="list-style-type: none">• Peer review of project reports <p>Assessment 2.2</p> <ul style="list-style-type: none">• In-class exam covering module II – Speech Synthesis



Teaching Week	Topics
11	<p>Speaker Identification</p> <p>Lecture</p> <ul style="list-style-type: none">• Overview in Speaker Identification• Discussion of relevant papers:<ul style="list-style-type: none">- Snyder, David, et al. <i>X-vectors: Robust DNN embeddings for speaker recognition.</i>, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018- Chung, et al., <i>VoxCeleb2: Deep Speaker Recognition.</i>, Proc. Interspeech, 2018- Li, Chao, et al., <i>Deep speaker: an end-to-end neural speaker embedding system.</i>, 2017 <p>Lab</p> <ul style="list-style-type: none">• Discussion on choosing a relevant paper to implement for the project• Start Project-3 work
12	<p>Speaker Identification</p> <p>Lecture</p> <ul style="list-style-type: none">• Speaker Identification – datasets and evaluation• Group discussion on the relevant papers <p>Lab</p> <ul style="list-style-type: none">• Continue project-3 work
13	<p>Speaker Identification</p> <p>Lecture</p> <ul style="list-style-type: none">• Speaker Identification – Deep learning models in Speaker Identification• Group discussion on the relevant papers <p>Lab</p> <ul style="list-style-type: none">• Continue project-3 work
14	<p>Speaker Identification</p> <p>Lecture</p> <ul style="list-style-type: none">• Speaker Identification – Few shots Speaker Identification <p>Lab</p> <ul style="list-style-type: none">• Preparation of presentation on project-3 work• Continue project-3 work
15	<p>Speaker Identification</p> <p>Assessment 3.1</p> <ul style="list-style-type: none">• Presentation of the projects by different groups <p>Lab</p> <ul style="list-style-type: none">• Peer review of project reports <p>Assessment 3.2</p> <ul style="list-style-type: none">• In-class exam covering module III – Speaker Identification