The course is focused on the study of the technologies that solve the problems associated with the development of advanced applications of spoken language. It focuses on new technologies based on deep learning and its application to automatic speech recognition. The final project gives students additional information about a particular topic, and also aims to help boost their own skills in the development of applications or in research.
## Study load

<table>
<thead>
<tr>
<th>Total learning time: 125h</th>
<th>Hours large group:</th>
<th>39h</th>
<th>31.20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours medium group:</td>
<td>0h</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>Hours small group:</td>
<td>0h</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>Guided activities:</td>
<td>0h</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td>Self study:</td>
<td>86h</td>
<td>68.80%</td>
<td></td>
</tr>
</tbody>
</table>
# 230623 - ST - Speech Technologies

## Content

| 1. Introduction to speech technology and speech processing | Learning time: 6h  
Theory classes: 6h |
| --- | --- |
| **Description:**  
Fundamentals of speech perception and speech production  
Speech signal: pdf, energy, zero crossing  
Speech spectrum: STFT, LPC, Spectrogram  
Features: LPC, Cepstrum, MFCC, Pitch  
Techniques: GMM, VQ | |

| 2. Classifiers | Learning time: 9h  
Theory classes: 9h |
| --- | --- |
| **Description:**  
Vector Quantizers, Gaussian Mixture Models, CART, Hidden Markov Models | |

| 3. Fundamentals of automatic speech recognition (ASR) | Learning time: 9h  
Theory classes: 9h |
| --- | --- |
| **Description:**  
Pattern matching. Dynamic time warping  
Hidden Markov models. Isolated word recognition  
Large vocabulary continuous ASR: Acoustic modeling, Language modeling, Search  
Toolkits | |

| 4. Speech synthesis | Learning time: 9h  
Theory classes: 9h |
| --- | --- |
| **Description:**  
Linguistic processing  
Prosody modeling  
Waveform generation  
Concatenation methods  
Statistical methods | |
230623 - ST - Speech Technologies

5. Deep Learning

<table>
<thead>
<tr>
<th>Description:</th>
<th>Learning time: 12h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 12h</td>
</tr>
</tbody>
</table>

Specific objectives:
- Deep learning techniques and their application to speech and language processing.

6. Other speech technologies

<table>
<thead>
<tr>
<th>Description:</th>
<th>Learning time: 6h</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theory classes: 6h</td>
</tr>
</tbody>
</table>

Specific objectives:
- Detection of speech and other sounds
- Speaker recognition and verification
- Speech-to-speech translation
- Neural Machine Translation

Qualification system

- First midterm exam: 20%
- Second midterm exam: 10%
- Assignments: 30%
- Research work: 40%
- Or final exam

Bibliography

Basic:
