320123 - TP - Speech Technology

Coordinating unit: 205 - ESEIAAT - Terrassa School of Industrial, Aerospace and Audiovisual Engineering
Teaching unit: 739 - TSC - Department of Signal Theory and Communications
Academic year: 2018
Degree: BACHELOR'S DEGREE IN AUDIOVISUAL SYSTEMS ENGINEERING (Syllabus 2009). (Teaching unit Optional)
ECTS credits: 6
Teaching languages: Catalan, Spanish

Teaching staff
Coordinator: Ignasi Esquerra
Others: Albino Nogueiras

Prior skills
It is recommended to have taken the courses in "Digital Audio Processing" and "Algorithms and Audiovisual Programming", and to have a certain knowledge in programming in Linux platforms.

Degree competences to which the subject contributes
Specific: CE26. (ENG) AUD: Coneixements i capacitats per aprofundir en tecnologies específiques de l'àmbit.

Teaching methodology
Learning with a teacher:
Autonomous learning:
b) Individual study, exercise solving and handouts preparation.

Learning objectives of the subject
Understanding the technologies used in human-computer interaction systems by voice. Students by the end of the course will learn several techniques of speech signal processing and will be able to put them into practice in speech recognition and speech synthesis systems.

Study load

<table>
<thead>
<tr>
<th>Total learning time: 150h</th>
<th>Hours large group: 30h</th>
<th>20.00%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours medium group: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Hours small group: 30h</td>
<td>20.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 90h</td>
<td>60.00%</td>
</tr>
</tbody>
</table>
### Content

<table>
<thead>
<tr>
<th>Topic</th>
<th>Learning time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. HUMAN-COMPUTER INTERACTION BY VOICE</strong></td>
<td>10h</td>
<td><strong>Description:</strong> Spoken language technologies. Historical review of Human-Computer interaction by voice. Applications and current research lines.</td>
</tr>
<tr>
<td><strong>2. DATABASES</strong></td>
<td>30h</td>
<td><strong>Description:</strong> Text and speech databases. Phonetics and linguistics. Production and perception of speech. Representation and analysis of speech signals.</td>
</tr>
<tr>
<td><strong>4. SPEECH RECOGNITION</strong></td>
<td>64h</td>
<td><strong>Description:</strong> Parametrization. Acoustic models. Hidden Markov Models. Language models.</td>
</tr>
</tbody>
</table>

### Qualification system

The final mark is the weighted sum of partial marks of course units. Each unit is assessed with several laboratory assignments or exams. None of the qualification elements has a weight over 25%.

Unit 1 (10%), Unit 2 (20%), Unit 3 (35%), Unit 4 (35%)
Bibliography

Complementary:
