230622 - DSAP - Digital Speech and Audio Processing

Coordinating unit: 230 - ETSETB - Barcelona School of Telecommunications Engineering
Teaching unit: 739 - TSC - Department of Signal Theory and Communications
Academic year: 2017
Degree: DEGREE IN TELECOMMUNICATIONS ENGINEERING (Syllabus 1992). (Teaching unit Optional)
MASTER’S DEGREE IN TELECOMMUNICATIONS ENGINEERING (Syllabus 2013). (Teaching unit Optional)
MASTER’S DEGREE IN INFORMATION AND COMMUNICATION TECHNOLOGIES (Syllabus 2009). (Teaching unit Optional)
ECTS credits: 5
Teaching languages: English

Teaching staff
Coordinator: Climent Nadeu
Others: Antonio Bonafonte
Javier Hernando

Opening hours
Timetable: Tuesday and Thursday from 10:00 to 13:00

Prior skills
Signal Processing

Requirements
Signal processing

Degree competences to which the subject contributes

Specific:
1. Ability to apply information theory methods, adaptive modulation and channel coding, as well as advanced techniques of digital signal processing to communication and audiovisual systems.

Transversal:
2. TEAMWORK: Being able to work in an interdisciplinary team, whether as a member or as a leader, with the aim of contributing to projects pragmatically and responsibly and making commitments in view of the resources that are available.
3. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.
4. FOREIGN LANGUAGE: Achieving a level of spoken and written proficiency in a foreign language, preferably English, that meets the needs of the profession and the labour market.

Teaching methodology
- Lectures (50%)
- Application classes (with Matlab or similar) (50%)
- Team work: project, presentation
- Individual work: preparation and completion (out classroom) of application activities

Learning objectives of the subject
Learning objectives of the subject
Understanding and being competent on a relevant set of concepts and techniques in the field of digital audio processing, and their application to problems arising from real applications. Especially, speech and music signals and applications will be considered.

Learning results:
Ability to digitally process, in an application-oriented context, audio and speech signals, in order to analyze, model, extract information from, clean, modify, and generate/synthesize them.

<table>
<thead>
<tr>
<th>Total learning time: 125h</th>
<th>Hours large group: 39h</th>
<th>31.20%</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: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Guided activities: 0h</td>
<td>0.00%</td>
</tr>
<tr>
<td></td>
<td>Self study: 86h</td>
<td>68.80%</td>
</tr>
</tbody>
</table>
# Content

<table>
<thead>
<tr>
<th>1. Introduction</th>
<th><strong>Learning time:</strong> 12h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>Theory classes: 6h</td>
</tr>
<tr>
<td></td>
<td>Self study : 6h</td>
</tr>
</tbody>
</table>

- Course presentation
- Audio diversity
- Characteristics of speech and music. Production model
- Hearing and auditory modeling
- The short-time Fourier transform

<table>
<thead>
<tr>
<th>2. Short-term analysis-synthesis of (cuasi)periodic signals</th>
<th><strong>Learning time:</strong> 12h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>Theory classes: 6h</td>
</tr>
<tr>
<td></td>
<td>Self study : 6h</td>
</tr>
</tbody>
</table>

- Filter-bank analysis/synthesis. The phase vocoder
- Filter-bank and spectrogram
- Time-scale and pitch modification
- QMF filters. MP3 coding.

<table>
<thead>
<tr>
<th>3. Modeling and representation of speech signals</th>
<th><strong>Learning time:</strong> 12h</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description:</strong></td>
<td>Theory classes: 6h</td>
</tr>
<tr>
<td></td>
<td>Self study : 6h</td>
</tr>
</tbody>
</table>

- Production-based all-pole modeling
- Pitch determination for speech and music
- LPC-based coding used in mobile telephony
### 4. Enhancement of speech and audio signals

**Learning time:** 12h  
Theory classes: 6h  
Self study: 6h

**Description:**  
Cancellation: echo, interference  
Denoising: spectral subtraction, Wiener-based filtering, wavelets  
Blind source separation: ICA, CASA, NMF

### 5. Multi-microphone audio processing

**Learning time:** 12h  
Theory classes: 6h  
Self study: 6h

**Description:**  
Room acoustics  
Array beamforming  
Acoustic source localization and tracking

### 6. Recognition and detection of audio and speech

**Learning time:** 12h  
Theory classes: 6h  
Self study: 6h

**Description:**  
6. Recognition and detection of audio and speech  
Pattern-matching approaches  
Audio activity detection  
Application to speech and speaker recognition
Projects realization and presentation

<table>
<thead>
<tr>
<th>Learning time: 54h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory classes: 3h</td>
</tr>
<tr>
<td>Self study: 51h</td>
</tr>
</tbody>
</table>

Description:
Design, implementation and test of a audio processing system for a specific application
Oral presentation of 1) Project proposal, and 2) Project realization

Qualification system

Attendance/participation in class (10%)
Tests (30%)
Project (50%)
Presentation (10%)

Bibliography

Basic:


Complementary:


Others resources:
Lecture slides
Practical work statements and programs

Audiovisual material

Slides
Slides used in lectures

Computer material

Codi programes
Software codes in Matlab or similar