



# Course guide

## 230022 - PAV - Audio and Speech Processing

Last modified: 30/05/2026

**Unit in charge:** Barcelona School of Telecommunications Engineering  
**Teaching unit:** 739 - TSC - Department of Signal Theory and Communications.

**Degree:** BACHELOR'S DEGREE IN TELECOMMUNICATIONS TECHNOLOGIES AND SERVICES ENGINEERING (Syllabus 2015). (Optional subject).

**Academic year:** 2026    **ECTS Credits:** 6.0    **Languages:** Spanish

### LECTURER

---

**Coordinating lecturer:** FRANCISCO JAVIER HERNANDO PERICAS - ALBINO NOGUEIRAS RODRIGUEZ

**Others:** Primer quadrimestre:  
FRANCISCO JAVIER HERNANDO PERICAS - 41  
ALBINO NOGUEIRAS RODRIGUEZ - 41

Segon quadrimestre:  
FRANCISCO JAVIER HERNANDO PERICAS - 21  
ALBINO NOGUEIRAS RODRIGUEZ - 21

### PRIOR SKILLS

---

Basic knowledge on signal and speech processing

### REQUIREMENTS

---

INTRODUCCIÓ AL PROCESSAMENT AUDIOVISUAL - Precorequisit

### DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

---

**Transversal:**

07 AAT N3. SELF-DIRECTED LEARNING - Level 3. Applying the knowledge gained in completing a task according to its relevance and importance. Deciding how to carry out a task, the amount of time to be devoted to it and the most suitable information sources.

### TEACHING METHODOLOGY

---

Expository classes  
Laboratory classes  
Group work (non-presential)  
Individual work (non-presential)  
Other activities  
Exercises  
Oral presentation  
Long Answer Tests (Control)  
Long Answer Tests (Final Exam)  
Laboratory practice

## LEARNING OBJECTIVES OF THE SUBJECT

The aim of the course is for the student to learn or deepen in the most common techniques for voice and audio processing as well as their use in problems of analysis, classification, synthesis or other types of processing.

Learning Outcome:

The subject contributes to the learning outcome specifically in the aspect of audio and voice information processing:

It has the capacity to build, operate and manage telecommunications services and applications, understood as capture systems, analog and digital processing, encoding, transport, representation, processing, storage, reproduction, management and presentation of audiovisual services and multimedia information.

He is familiar with the analysis, specification, operation and maintenance of television and audio systems, equipment, headers and installations, both in fixed and mobile environments.

It is able to carry out projects of premises and facilities for the production and recording of audio signals.

It has the ability to create, encode, manage, disseminate and distribute multimedia content, meeting the criteria of usability and accessibility of audiovisual, broadcast and interactive services.

In teamwork, plan and agree on goals, operating rules, responsibilities, agenda, and work review procedure.

Carry out the tasks based on the basic guidelines given by the teacher, deciding the time and resources needed. Assess your own strengths and weaknesses and act accordingly.

## STUDY LOAD

Type	Hours	Percentage
Hours large group	39,0	26.00
Self study	85,0	56.67
Hours small group	26,0	17.33

**Total learning time:** 150 h

## CONTENTS

### Topic 1. Introduction to speech and written language processing

**Description:**

Speech  
Written Language  
Speech and Written Language Technologies

**Full-or-part-time:** 2h 30m

Theory classes: 2h  
Self study : 0h 30m

### Topic 2. The speech signal

**Description:**

Human Language  
Speech Production  
Definition and Analysis of Speech Characteristics

**Full-or-part-time:** 6h 05m

Theory classes: 2h  
Self study : 4h 05m



### Topic 3. Representation of the voice signal. Dynamic programming

**Description:**

Pitch estimation  
Homomorphic processing: the cepstrum  
Mel-spectrogram and mel-cepstrum  
Dynamic programming for isolated word recognition

**Full-or-part-time:** 21h 20m

Theory classes: 5h 20m

Self study : 16h

### Topic 4. Statistical Modeling. Viterbi

**Description:**

Gaussian Mixture Models. Voice classification  
Hidden Markov Models  
Viterbi algorithm. Speech recognition  
N-grams. Language models

**Full-or-part-time:** 25h 25m

Theory classes: 8h

Self study : 17h 25m

### Topic 5. Recurrent Neural Networks. Attention

**Description:**

Recurrent Neural Networks. Language Models  
Attention  
Sequence-to-Sequence Architectures. Translation. Text-to-Speech conversion

**Full-or-part-time:** 16h

Theory classes: 8h

Self study : 8h

### Topic 6. Transformer

**Description:**

Positional embedding  
Multihead and cross-attention  
Transformer  
Neural speech recognition

**Full-or-part-time:** 16h

Theory classes: 8h

Self study : 8h



### Topic 7. Self-supervised learning and foundational models

**Description:**

BERT, ChatGPT, HuBERT, mBART  
Speech language models

**Full-or-part-time:** 25h

Theory classes: 12h  
Self study : 13h

### (ENG) Projecte: desenvolupament de sistema de processat de veu i/o àudio

**Full-or-part-time:** 30h

Laboratory classes: 13h  
Self study : 17h

## ACTIVITIES

(ENG) Exercicis

(ENG) Proves de resposta llarga (Control)

(ENG) Pràctica de laboratori

(ENG) Altres activitats

(ENG) Altres activitats

(ENG) Proves de resposta llarga (Examen Final)

## GRADING SYSTEM

---

Partial control (CNT): 25%  
Laboratory practices (LAB): 35%  
Final exam (FNL): 40%

In the theory section, with a combined weight of 65% of the grade, the highest between the average of the two exams (0.5 CNT + 0.5 FNL) and the final exam grade (FNL) is applied.

In this subject the generic competences will be evaluated:

- Teamwork (Intermediate Level)
- Autonomous learning (Intermediate level)

## EXAMINATION RULES.

---

## BIBLIOGRAPHY

---

### Basic:

- Huang, X.; Acero, A.; Hon, H-W. Spoken language processing: a guide to theory, algorithm and system development. Upper Saddle River: Prentice Hall, 2001. ISBN 0130226165.
- Jurafsky, D; Martin, J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models [on line]. 3rd ed, Online manuscript released. Stanford, 2026 [Consultation: 01/06/2026]. Available on: <https://web.stanford.edu/~jurafsky/slp3>.