



# Guia docent

## 230022 - PAV - Processament d'Àudio i Veu

Última modificació: 10/06/2025

**Unitat responsable:** Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona  
**Unitat que imparteix:** 739 - TSC - Departament de Teoria del Senyal i Comunicacions.

**Titulació:** GRAU EN ENGINYERIA DE TECNOLOGIES I SERVEIS DE TELECOMUNICACIÓ (Pla 2015). (Assignatura optativa).

**Curs:** 2025      **Crèdits ECTS:** 6.0      **Idiomes:** Castellà

### PROFESSORAT

---

**Professorat responsable:** FRANCISCO JAVIER HERNANDO PERICAS - ALBINO NOGUEIRAS RODRIGUEZ

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

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

### CAPACITATS PRÈVIES

---

Coneixements bàsics de processament d'àudio  
Es recomana cursar o haver cursat el seminari de grau Aprenentatge automàtic

### REQUISITS

---

INTRODUCCIÓ AL PROCESSAMENT AUDIOVISUAL - Precorequisit

### COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

---

**Transversals:**

07 AAT N3. APRENTATGE AUTÒNOM - Nivell 3: Aplicar els coneixements assolits a la realització d'una tasca en funció de la pertinència i la importància, decidint la manera de dur-la a terme i el temps que cal dedicar-hi i seleccionant-ne les fonts d'informació més adequades.

### METODOLOGIES DOCENTS

---

Classes expositives  
Classes laboratori  
Treball en grup (no presencial)  
Treball individual (no presencial)  
Altres activitats  
Exercicis  
Presentació oral  
Proves de resposta llarga (Control)  
Proves de resposta llarga (Examen Final)  
Pràctica de laboratori

## OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

L'objectiu del curs és que l'alumne aprengui o aprofundeixi en les tècniques més habituals per al processat de veu i d'àudio així com la seva utilització en problemes d'anàlisi, classificació, síntesi o altre tipus de processat.

Resultat de l'aprenentatge:

L'assignatura contribueix al resultat de l'aprenentatge específicament en l'aspecte de processat d'informació d'àudio i veu:

Té capacitat de construir, explotar i gestionar serveis i aplicacions de telecomunicacions, enteses aquestes com sistemes de captació, tractament analògic i digital, codificació, transport, representació, processat, emmagatzematge, reproducció, gestió i presentació de serveis audiovisuals i informació multimèdia.

Esta familiaritzat amb l'anàlisi, especificació, operació i manteniment de sistemes, equips, capçaleres i instal·lacions de televisió i àudio, tant en entorns fixos com mòbils.

És capaç de realitzar projectes de locals i instal·lacions destinats a la producció i enregistrament de senyals d'àudio.

Té capacitat per a crear, codificar, gestionar, difondre i distribuir continguts multimèdia, atenent a criteris de usabilitat i accessibilitat dels serveis audiovisuals, de difusió i interactius.

En els treballs en equip, planifica i acorda els objectius, les regles de funcionament, les responsabilitats, l'agenda i el procediment de revisió del treball.

Porta a terme les tasques a partir de les orientacions bàsiques donades pel professorat, decidint el temps i els recursos necessaris. Avalua les pròpies fortaleces i debilitats i actua en conseqüència.

## HORES TOTALES DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores aprenentatge autònom	85,0	56.67
Hores grup gran	39,0	26.00
Hores grup petit	26,0	17.33

**Dedicació total:** 150 h

## CONTINGUTS

### Tema 1. Introducció al processat de veu i àudio

**Dedicació:** 2h 30m

Grup gran/Teoria: 2h

Aprenentatge autònom: 0h 30m

### Tema 2. El senyal de veu

**Descripció:**

El llenguatge humà

Producció de la parla

Definició i anàlisi de característiques del parla

**Dedicació:** 6h 05m

Grup gran/Teoria: 2h

Aprenentatge autònom: 4h 05m



### Tema 3. Representacions de senyals de veu i àudio

**Descripció:**

Representacions basades en STFT (short-time Fourier transform)  
Representacions basades en LPC (predicció lineal)  
Processat homomòrfic: el cepstrum Estimació pitch

**Dedicació:** 21h 20m

Grup gran/Teoria: 5h 20m

Aprenentatge autònom: 16h

### Tema 4. Robustesa respecte a l'entorn acústic

**Descripció:**

Eco i reverberació  
Arrays de micròfons  
Substracció espectral

**Dedicació:** 25h 25m

Grup gran/Teoria: 8h

Aprenentatge autònom: 17h 25m

### Tema 5. Fonaments de la classificació de veu i àudio

**Descripció:**

Nocions de classificadores: regressió lineal, CART (clasificacions and regression trees)  
Clasificadors bayesianos.  
GMM (gaussian mixture models)  
HMM (Hidden Markov Models)  
Detecció de veu  
Detecció d'esdeveniments acústics  
Reconeixement del parlant

**Dedicació:** 16h

Grup gran/Teoria: 8h

Aprenentatge autònom: 8h

### Tema 6. Reconeixement de la parla

**Descripció:**

Modelatge acústic  
Modelatge del llenguatge

**Dedicació:** 16h

Grup gran/Teoria: 8h

Aprenentatge autònom: 8h



### Tema 7. Síntesi de la parla

**Descripció:**

Processat del text  
Descripció i modelatge de la prosodia  
Generació paramètrica.  
Model de formantes.  
Generació de veu mitjançant concatenació. Mètodes PSOLA i HSM.  
Generació de veu mitjançant selecció d'unitats.  
Generació paramètrica mitjançant HMM.

**Dedicació:** 25h

Grup gran/Teoria: 12h

Aprenentatge autònom: 13h

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

**Descripció:**

En grups s'haurà de dissenyar i desenvolupar un sistema de processat de veu i àudio.

Exemple: millora d'un sistema de reconeixement de la parla, intervenint en diferents aspectes. Tots els projectes es presenten en l'aula a professors i estudiants.

**Dedicació:** 30h

Grup petit/Laboratori: 13h

Aprenentatge autònom: 17h

## ACTIVITATS

### Exercicis

**Descripció:**

Els alumnes realitzen individualment 5 treballs

### Proves de resposta llarga (Control)

**Descripció:**

Exàmens parcials

### Pràctica de laboratori

**Descripció:**

Projecte: desenvolupament de sistema de processat de veu i/o àudio

### Altres activitats

**Descripció:**

Projecte: desenvolupament de sistema de processat de veu i/o àudio



#### Altres activitats

**Descripció:**

Projecte: desenvolupament de sistema de processat de veu i/o àudio

#### Proves de resposta llarga (Examen Final)

**Descripció:**

Examen final

## SISTEMA DE QUALIFICACIÓ

---

Control parcial (CNT): 32.5%  
Pràctiques laboratori (LAB): 35%  
Examen final (FNL): 32.5%

A la part de teoria, amb un pes del 65% de la nota, s'aplica la major entre la mitja dels dos exàmens (0.5 CNT + 0.5 FNL) i la nota de l'exàmen final (FNL).

En aquesta assignatura s'avaluaran les competències genèriques:

- Treball en equip (Nivell Mitjà)
- Aprenentatge autònom (Nivell Mitjà)

## NORMES PER A LA REALITZACIÓ DE LES PROVES.

---

## BIBLIOGRAFIA

---

**Bàsica:**

- 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.