

Guía docente

230022 - PAV - Procesado de Voz y Lenguaje Escrito con Inteligencia Artificial

Última modificación: 30/05/2026

Unidad responsable: Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona

Unidad que imparte: 739 - TSC - Departamento de Teoría de la Señal y Comunicaciones.

Titulación: GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN (Plan 2015). (Asignatura optativa).

Curso: 2026

Créditos ECTS: 6.0

Idiomas: Castellano

PROFESORADO

Profesorado responsable: ALBINO NOGUEIRAS RODRIGUEZ

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

CAPACIDADES PREVIAS

Conocimientos básicos de procesado de señal y voz

REQUISITOS

INTRODUCCIÓ AL PROCESSAMENT AUDIOVISUAL - Precorequisit

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Transversales:

07 AAT N3. APRENDIZAJE AUTÓNOMO - Nivel 3: Aplicar los conocimientos alcanzados en la realización de una tarea en función de la pertinencia y la importancia, decidiendo la manera de llevarla a cabo y el tiempo que es necesario dedicarle y seleccionando las fuentes de información más adecuadas.

METODOLOGÍAS DOCENTES

Clases expositivas
Clases laboratorio
Trabajo en grupo (no presencial)
Trabajo individual (no presencial)
Otras actividades
Ejercicios
Presentación oral
Pruebas de respuesta larga (Control)
Pruebas de respuesta larga (Examen Final)
Práctica de laboratorio

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

El objetivo del curso es que el alumno aprenda o profundice en las técnicas más habituales para el procesado de voz y audio así como su utilización en problemas de análisis, clasificación, síntesis u otro tipo de procesado.

Resultado del aprendizaje:

La asignatura contribuye al resultado del aprendizaje específicamente en el aspecto de procesado de información de audio y voz:

Tiene capacidad de construir, explotar y gestionar servicios y aplicaciones de telecomunicaciones, entendidas éstas como sistemas de captación, tratamiento analógico y digital, codificación, transporte, representación, procesado, almacenamiento, reproducción, gestión y presentación de servicios audiovisuales e información multimedia.

Está familiarizado con el análisis, especificación, operación y mantenimiento de sistemas, equipos, cabeceras e instalaciones de televisión y audio, tanto en entornos fijos como móviles.

Es capaz de realizar proyectos de locales e instalaciones destinados a la producción y grabación de señales de audio.

Tiene capacidad para crear, codificar, gestionar, difundir y distribuir contenidos multimedia, atendiendo a criterios de usabilidad y accesibilidad de los servicios audiovisuales, de difusión e interactivos.

En los trabajos en equipo, planifica y acuerda los objetivos, reglas de funcionamiento, responsabilidades, agenda y procedimiento de revisión del trabajo.

Lleva a cabo las tareas a partir de las orientaciones básicas dadas por el profesorado, decidiendo el tiempo y los recursos necesarios.

Evalúa las propias fortalezas y debilidades y actúa en consecuencia.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTE

Tipo	Horas	Porcentaje
Horas aprendizaje autónomo	85,0	56.67
Horas grupo grande	39,0	26.00
Horas grupo pequeño	26,0	17.33

Dedicación total: 150 h

CONTENIDOS

Tema 1. Introducción al procesamiento de la voz y el lenguaje escrito

Descripción:

La voz

El lenguaje escrito

Tecnologías de la voz y el lenguaje escrito

Dedicación: 2h 30m

Grupo grande/Teoría: 2h

Aprendizaje autónomo: 0h 30m

Tema 2. La señal de voz

Descripción:

El lenguaje humano

Producción del habla

Definición y análisis de características del habla

Dedicación: 6h 05m

Grupo grande/Teoría: 2h

Aprendizaje autónomo: 4h 05m

Tema 3. Representación de la señal de voz. Programación dinámica

Descripción:

Estimación de pitch

Procesado homomórfico: el cepstrum

Mel-espectrograma y mel-cepstrum

Programación dinámica para reconocimiento de palabras aisladas

Dedicación: 21h 20m

Grupo grande/Teoría: 5h 20m

Aprendizaje autónomo: 16h

Tema 4. Modelado estadístico. Viterbi

Descripción:

Modelos de mezclas de gaussianas. Clasificación de voz

Modelos ocultos de Markov

Algoritmo de Viterbi. Reconocimiento de voz

N-gramas. Modelos de lenguaje

Dedicación: 25h 25m

Grupo grande/Teoría: 8h

Aprendizaje autónomo: 17h 25m

Tema 5. Redes Neuronales Recurrentes. Atención

Descripción:

Redes Neuronales Recurrentes. Modelos de lenguaje

Atención

Arquitecturas secuencia en secuencia. Traducción. Conversión texto a voz

Dedicación: 16h

Grupo grande/Teoría: 8h

Aprendizaje autónomo: 8h

Tema 6. Transformer

Descripción:

Embeddings posicionales

Atención multi-head y cruzada

Transformer

Reconocimiento neuronal del habla

Dedicación: 16h

Grupo grande/Teoría: 8h

Aprendizaje autónomo: 8h



Tema 7. Aprendizaje autosupervisado y modelos fundacionales

Descripción:

BERT, ChatGPT, HuBERT, mBART
Modelos de lenguaje de voz

Dedicación: 25h

Grupo grande/Teoría: 12h
Aprendizaje autónomo: 13h

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

Dedicación: 30h

Grupo pequeño/Laboratorio: 13h
Aprendizaje autónomo: 17h

ACTIVIDADES

(CAST) Exercicis

(CAST) Proves de resposta llarga (Control)

(CAST) Pràctica de laboratori

(CAST) Altres activitats

(CAST) Altres activitats

(CAST) Proves de resposta llarga (Examen Final)

SISTEMA DE CALIFICACIÓN

Cualquier acto de fraude académico, plagio o uso o mera tenencia al alcance de medios no autorizados en cualquier actividad de evaluación comportará la calificación de cero (0) en la prueba o entrega afectada. Además, de acuerdo con la normativa de la Universidad, la posible derivación de los hechos para la apertura de un expediente disciplinario implicará que la asignatura quede en el estado provisional de "pendiente de evaluación" hasta la resolución del expediente. La gestión de estas incidencias se lleva a cabo de acuerdo con el [Marco de actuación para la integridad académica en la evaluación de la UPC](#).

Control parcial (CNT): 25%
Prácticas laboratorio (LAB): 35%
Examen final (FNL): 40%

En la parte de teoría, con un peso conjunto del 65% de la nota, se aplica la mayor entre la media de los dos exámenes (0.5 CNT + 0.5 FNL) y la nota del examen final (FNL).

En esta asignatura se evaluarán las competencias genéricas:

- Trabajo en equipo (Nivel Medio)
- Aprendizaje autónomo (Nivel Medio)

NORMAS PARA LA REALIZACIÓN DE LAS PRUEBAS.

BIBLIOGRAFÍA

Básica:

- 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 [en línea]. 3rd ed, Online manuscript released. Stanford, 2026 [Consulta: 01/06/2026]. Disponible a: <https://web.stanford.edu/~jurafsky/slp3>.