



FICHA IDENTIFICATIVA

DATOS DE LA ASIGNATURA

Código: 34907

Nombre: Procesado digital de audio y voz

Ciclo: Grado

Créditos ECTS: 6

Curso académico: 2026-27

TITULACIONES

Titulación	Centro	Curso	Periodo
1403 - Grado en Ingeniería Telemática	Escola Tècnica Superior d'Enginyeria	4	Segundo cuatrimestre

MATERIAS

Titulación	Materia	Carácter
1403 - Grado en Ingeniería Telemática	Optatividad	OPTATIVA

COORDINACIÓN

COBOS SERRANO MAXIMO

RESUMEN

La asignatura *¿Procesado Digital de Audio y Voz¿* es una asignatura de cuarto curso que forma parte de la oferta en optatividad del Grado en Ingeniería Telemática. La asignatura complementa los contenidos vistos en otras asignaturas del Grado como *¿Procesado Digital de la Señal¿*, *¿Señales y Sistemas Lineales¿* y *¿Fundamentos Matemáticos de las Comunicaciones¿*, ofreciendo una visión aplicada de los conceptos estudiados a lo largo de estas materias. Así, los temas abarcados por esta asignatura están orientados a la aplicación del tratamiento digital de señales y el aprendizaje automático en el campo del procesado de audio y de voz.

La asignatura justifica la importancia del procesado digital de señales de audio en los sistemas multimedia actuales, revisando brevemente algunos conceptos básicos estudiados en cursos anteriores. La estructuración de los contenidos sigue un enfoque en el que se comienza introduciendo los conceptos más teóricos en la primera parte del curso relacionados con el procesado digital de señales, dando paso a otros temas más enfocados al aprendizaje máquina, los cuales han adquirido una gran importancia en la actualidad. Así, se parte de una revisión de conceptos básicos de procesado de señal, introduciendo varias representaciones de las señales de audio, poniendo especial atención a las representaciones tiempo-frecuencia. La parte teórica se complementa con el estudio del sistema auditivo humano y su impacto en



el diseño de sistemas prácticos de codificación de audio con pérdidas, los fundamentos del sistema fonador y los modelos fuente/filtro utilizados en codificación de voz.

La segunda parte del curso introduce los fundamentos de los sistemas de audio basados en aprendizaje máquina, introduciendo el uso de características utilizadas ampliamente en problemas de audio y su combinación con clasificadores binarios clásicos como las máquinas de vectores soporte o regresión logística. Por último se introduce al alumno a los sistemas de aprendizaje profundo actuales.

Nota: en esta Guía, donde dice estudiante, se entiende también la estudiante, así como alumno también alumna, profesor también profesora, o viceversa.

CONOCIMIENTOS PREVIOS

RELACIÓN CON OTRAS ASIGNATURAS DE LA MISMA TITULACIÓN

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

OTROS TIPOS DE REQUISITOS

COMPETENCIAS / RESULTADOS DE APRENDIZAJE

DESCRIPCIÓN DE CONTENIDOS

0. Introducción

Visión global del procesado digital de señal de audio, imágenes y vídeo. El procesado de señal en sistemas multimedia.

1. Revisión de procesado de señal

Descripción de contenidos (Castellano):

Introducción. Revisión de convolución y filtrado. Teoría del muestreo. Procesado de señal en tiempo discreto. DFT y DTFT. Procesos aleatorios.



2. Audio y principios psicoacústicos

Descripción de contenidos (Castellano):

Introducción. Fundamentos de acústica. Audición y percepción. Sonoridad y bandas críticas.

3. Introducción al aprendizaje automático en audio

Procesado de señal e inteligencia artificial en audio. Evolución e historia de los sistemas de audio basados en aprendizaje automático. Retos relacionados con el uso de las técnicas de aprendizaje automático en audio. Aplicaciones actuales.

4. Representaciones de audio para aprendizaje automático

Representaciones de audio tiempo-frecuencia: STFT, Espectrograma Mel, Bancos de filtros Gammatone, Chromagrama. Características clásicas: temporales, espectrales, MFCCs. Representaciones compactas y reducción de dimensionalidad: NMF, PCA.

5. Técnicas de aprendizaje automático aplicadas a audio y voz

Introducción al aprendizaje supervisado y no supervisado. Clasificadores binarios lineales: SVM, Regresión logística. Clasificadores no lineales: k-NN, modelos Kernel. Modelos de mezclas de Gaussianas (GMM), k-Means. Introducción a las redes neuronales y el aprendizaje profundo.

VOLUMEN DE TRABAJO (HORAS)

ACTIVIDADES PRESENCIALES

Actividad	Horas
Teoría	30,00
Prácticas en aula	10,00
Laboratorio	20,00
Total horas	60,00

ACTIVIDADES NO PRESENCIALES

Actividad	Horas
Asistencia a otras actividades	0,00
Elaboración de trabajos individuales o en grupo	30,00
Estudio y trabajo autónomo	0,00
Preparación de clases	25,00
Preparación de actividades de evaluación	20,00
Resolución de casos prácticos	15,00
Total horas	90,00



METODOLOGÍA DOCENTE

1.- Trabajo presencial formado por:

a) Clases de teoría, las cuáles consistirán en la presentación y explicación básica de la materia correspondiente. Se propondrán actividades de corta duración, las cuales exigirán la intervención de los alumnos con el objetivo de confirmar la comprensión de la teoría expuesta.

b) Clases de ejercicios y seminarios prácticos, diseñadas para resolver problemas de mayor envergadura o bien temporal o bien conceptual.

c) Clases de laboratorio, pensadas para comprobar experimentalmente algunas de las cuestiones más relevantes vistas en las clases de teoría.

2.- Trabajo no presencial formado por:

a) Preparación de cuestionarios de tema a realizar tras finalizar cada unidad didáctica.

b) Preparación de las prácticas de laboratorio, para las que el alumno deberá haber leído y asimilado el contenido del boletín de prácticas, así como haber repasado la teoría relevante.



Tutorías individuales y/o colectivas. Se establecen unas determinadas horas de tutorías no programadas por semana a las que los alumnos podrán asistir para aclarar sus dudas (presencialmente o en modo on-line).

en modo on-line).

EVALUACIÓN

La evaluación de la asignatura se basará en un sistema de evaluación continua que considera el trabajo del alumno a lo largo del curso, complementado con una prueba final que medirá el conocimiento global adquirido.

Los componentes y su ponderación son los siguientes:

- Cuestionarios por tema (realizados tras la finalización de cada unidad didáctica): 45%
- Prueba práctica final (que evaluará los conocimientos obtenidos en las sesiones de laboratorio): 20%
- Cuestionario global final: 25%
- Asistencia, atención y participación en clase: 10%

La copia o plagio manifiesto de cualquier actividad que forma parte de la evaluación supondrá la imposibilidad de superar la asignatura, sometiéndose seguidamente a los procedimientos disciplinarios oportunos indicados en el PROTOCOLO DE

ACTUACIÓN ANTE PRÁCTICAS FRAUDULENTAS EN LA UNIVERSITAT DE VALÈNCIA (ACGUV 123/2020).

BIBLIOGRAFÍA

- Aurélien Géron, Hands-on Machine Learning with Scikit-Learn, Keras & Tensorflow, OReilly, 2020.
- Virtanen, Plumbley, Ellis, (eds): Computational Analysis of Sound Scenes and Events, Springer, 2018.



- Giannakopoulos, T. And Pikrakis, A. Introduction to Audio Analysis, Elsevier, 2014.
- Zölzer, Udo., Digital Audio Signal Processing, 2nd edition, Wiley, 2008. ISBN: 0470997850
- Pulkki, V., Karjalainen, M. Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics. Wiley (2015). ISBN: 978-1-118-86654-2
- -Smith III, Julius O., Spectral Audio Signal Processing, W3K Publishing, 2011. 978-0974560731
- -Bosi, M. and Goldberg, Richard E., Introduction to Digital Audio Coding and Standards, Kluwer Academic Publishers, 2003. ISBN: 978-1402073571