

**FITXA IDENTIFICATIVA****DADES DE L'ASSIGNATURA**

Codi: 36439
Nom: Anàlisi i processament d'àudio i veu
Cicle: Grau
Crèdits ECTS: 4,5
Curs acadèmic: 2025-26

TITULACIONS

Titulació	Centre	Curs	Període
1406 - Grau en Ciència de Dades	Escola Tècnica Superior d'Enginyeria	4	Segon quadrimestre

MATÈRIES

Titulació	Matèria	Caràcter
1406 - Grau en Ciència de Dades	Anàlisi i processament d'àudio i veu	OPTATIVA

COORDINACIÓ

SEGURA GARCIA JAUME

ROMERO GOMEZ VERONICA

GIMENEZ PASTOR ADRIA

RESUM

L'assignatura Anàlisi i processament d'Àudio i Veu és una assignatura de quart curs que forma part de la oferta en optativitat del Grau en Ciència de Dades. L'assignatura complementa els continguts vistos en altres assignatures del Grau com, Senyals i Sistemes, Aprenentatge màquina i Processat del llenguatge natural, oferint una visió aplicada dels conceptes estudiats al llarg d'aquestes matèries. Així, els temes abastats per aquesta assignatura estan orientats a l'aplicació del tractament digital de senyals en el camp del processat d'àudio i de veu. L'assignatura justifica la importància del processat digital de senyals d'àudio en els sistemes multimèdia i de processat massiu de dades actuals, revisant breument alguns conceptes bàsics estudiats en cursos anteriors. L'estructuració dels continguts segueix un enfocament en el que es comença introduint els conceptes més teòrics en la primera part del curs, donant pas a temes que utilitzen eixos conceptes en sistemes pràctics actuals. Així, es parteix d'una revisió de conceptes bàsics de processat de senyal, així com altres conceptes rellevants en àudio com el mostreig, la quantificació, les tècniques d'anàlisi/síntesi en temps curt i mètodes de processat estadístic. Es continua amb l'estudi del sistema auditiu humà i el seu impacte en el disseny de sistemes pràctics de codificació d'àudio amb pèrdues, els fonaments del sistema fonador i els models font/filtre utilitzats en codificació de veu. A més, s'introdueixen conceptes com filtrat òptim, predicció lineal o anàlisi sub-banda, que són introduïts des d'un punt de vista aplicat. La part teòrica es complementa amb els fonaments dels sistemes



de reconeixement de veu basats en models ocults de Markov i models de llenguatge. A més, s'estudien tècniques d'aprenentatge màquina amb aplicació al reconeixement de locutors i síntesi de veu. L'objectiu de l'assignatura és proporcionar a l'estudiant una visió aplicada dels conceptes estudiats al llarg del Grau, facilitant la comprensió de termes abstractes mitjançant l'estudi d'aplicacions concretes. Així, pretenem que l'estudiant obtinga una visió pràctica de l'estudi del processat i caracterització de la senyal veu, models de llenguatge, aprenentatge màquina, reconeixement i síntesi de veu.

CONEXEMENTS PREVIS

RELACIÓ AMB ALTRES ASSIGNATURES DE LA MATEIXA TITULACIÓ

No s'ha especificat restriccions de matrícula amb altres assignatures del pla d'estudis.

ALTRES TIPUS DE REQUISITS

COMPETÈNCIES / RESULTATS D' APRENENTATGE

-

(CB5) Que els estudiants hagen desenvolupat aquelles habilitats d'aprenentatge necessàries per a emprendre estudis posteriors amb un alt grau d'autonomia.

(CE03) Capacitat per resoldre problemes de classificació, modelització, segmentació i predicció a partir d'un conjunt de dades.

(CE10) Capacitat per processar senyals de forma digital extraient informació d'elles.

(CG01) Coneixement de matèries bàsiques i tecnologies, que li capacite per a l'aprenentatge de nous mètodes i tecnologies, així com que li dote d'una gran versatilitat per a adaptar-se a noves situacions.

(CT01) Ser capaços d'accedir a eines d'informació (bibliogràfiques) i d'utilitzar-les apropiadament en el desenvolupament de les seves tasques quotidianes.

DESCRIPCIÓ DE CONTINGUTS

1. Introducció.

Visió global del processat digital de senyal d'àudio, imatges i vídeo. El processat de senyal en sistemes multimèdia.

Introducció. Revisió de convolució i filtrat. Teoria del mostreig. Processat de senyal en temps discret. DFT i



2. Revisió de processat de senyal

DTFT. Processos aleatoris.

3. Anàlisi i síntesi d'àudio.

Introducció. Short-Time Fourier Transform. Bancs de filtres. Overlap-add.

4. Codificació.

Introducció. Predicció lineal. Codificació sub-banda i transformades. MPEG Audio

5. Introducció al reconeixement de veu.

Definició del problema. Evaluació dels sistemes. Clasificació de problemes. Historia del reconeixement de veu.

6. Reconeixement de veu.

Aproximació estadística. Models Ocults de Markov. Modelat de llenguatge. Decodificació.

7. Reconeixement de veu amb xarxes neuronals.

Recurrent Neural Networks, Long short-term memory, Connectionist Temporal Classification, Convolutional Neural Networks, RAH amb xarxes neuronals.

VOLUM DE TREBALL (HORES)

ACTIVITATS PRESENCIALS

Activitat	Hores
Teoria	22,00
Pràctiques a l'aula	8,00
Laboratori	15,00
Total hores	45,00

ACTIVITATS NO PRESENCIALS

Activitat	Hores
Assistència a altres activitats	2,00
Elaboració de treballs individuals o en grup	42,50



Estudi i treball autònom	0,00
Preparació de classes	6,00
Preparació d'activitats d'avaluació	2,00
Resolució de casos pràctics	15,00
Total hores	67,50

METODOLOGIA DOCENT

MD1 - Activitats teòriques. Desenvolupament expositiu de la matèria amb la participació de l'estudiant en la resolució de qüestions puntuals. Realització de qüestionaris individuals d'avaluació.

MD2 - Activitats pràctiques. Aprenentatge mitjançant resolució de problemes, exercicis i casos d'estudi a través dels quals s'adquireixen competències sobre els diferents aspectes de la matèria.

MD4 - Treballs en laboratori i/o aula amb ordinador. Aprenentatge mitjançant la realització d'activitats desenvolupades de forma individual o en grups reduïts i dutes a terme en laboratoris i/o aules d'ordinadors.

1) Treball presencial format per:

a) Classes de teoria, les quals consistiran en la presentació i explicació bàsica de la matèria corresponent. Es proposaran activitats de curta duració, les quals exigiran la intervenció de l'alumnat amb l'objectiu de confirmar la comprensió de la teoria exposada.

b) Classes d'exercicis, dissenyades per resoldre problemes de major envergadura o bé temporal o bé conceptual.

c) Classes de laboratori, pensades per comprovar experimentalment algunes de les qüestions més rellevants vistes en les classes de teoria.

2) Treball no presencial format per:

a) Resolució i presentació d'exercicis. Es tracta de resoldre els butlletins d'exercicis proposats pel professorat i/o l'exposició en públic de la resolució d'alguns d'ells.

b) Preparació dels exàmens.

c) Preparació de les pràctiques de laboratori, per a les que l'alumnat haurà d'haver llegit i assimilat el contingut del butlletí de pràctiques, així com haver repassat la teoria relevant.

Tutories individuals i/o col·lectives. S'estableixen unes determinades hores de tutories no programades per setmana a les que l'alumnat podran assistir per aclarar els seus dubtes

AVALUACIÓ

SE1 - Prova objectiva, consistent en dos exàmens que consten tant de qüestions teòric-pràctiques com de problemes. 30%+30%

SE2 - Avaluació de les activitats pràctiques a partir de l'elaboració de treballs/memòries i/o exposicions orals. 35%

SE3 - Avaluació contínua de cada alumne, basada en la participació i grau d'implicació de l'alumne en el procés d'ensenyament-aprenentatge, tenint en compte l'assistència regular a les activitats presencials previstes i la resolució de qüestions i problemes proposats periòdicament. 5%



El mecanisme d'avaluació és el que es podria anomenar tradicional adaptat, que no arriba a ser una avaluació contínua completa. Es tenen en compte els següents ítems i valoracions:

Valoració de l'assistència i la participació (5% de la nota final). Nota no recuperable, es conserva la nota obtinguda en segona convocatòria.

Dos examens parcials (30% de la nota final cadascú).

Assistència i realització de les pràctiques (20% de la nota final). Nota no recuperable, es conserva la nota obtinguda en segona convocatòria.

Resolució d'exercicis proposats (15% de la nota final). Nota no recuperable, es conserva la nota obtinguda en segona convocatòria.

Es considera que un alumne assisteix regularment a classe quan no ha faltat a més del 25% de les classes donades fins a la meitat del temps d'impartició de l'assignatura.

Per als alumnes que no puguem assistir regularment a classe, s'ofereix un model alternatiu en el que la valoració de l'assistència i participació es substituiran per treballs addicionals amb un percentatge total equivalent. Els alumnes que vulguen acollir-se a aquesta opció hauran de comunicar-ho al professor durant les 3 primeres setmanes del curs.

Els mínims requerits per superar l'assignatura seran l'equivalent a un 3.5 sobre 10 en els examens parcial. La resta d'ítems avaluable no estan sotmesos a mínim.

Tant en primera com en segona convocatòria, aquells estudiants que no superen el 5 mitjançant les activitats d'avaluació contínua, podran elegir entre dues opcions:

- Examen final (60%). La nota corresponent als examens parcials es recupera amb un únic examen final.
- Examens parcials (30%/30%). Es guarda la nota del examen parcial superior a 3.5 i es recupera la part suspesa mitjançant examen.

La còpia o plagi manifest de qualsevol activitat que forma part de l'avaluació suposarà la impossibilitat de superar l'assignatura, sotmetent-se seguidament als procediments disciplinaris oportuns indicats en el PROTOCOL D'ACTUACIÓ DAVANT PRÀCTIQUES FRAUDULENTES A LA UNIVERSITAT DE VALÈNCIA (ACGUV 123/2020).

BIBLIOGRAFIA

- Wendy Holmes. Speech synthesis and recognition [2nd ed]
- Sadaoki Furui. Digital Speech Processing, Synthesis and Recognition [2 ed.]
- José Unpingco. Python for Signal Processing. Featuring IPython Notebooks. Springer International, 2014 ISBN 978-3-319-01341-1
- Statistical Methods for Speech Recognition. Jelinek. The MIT Press. 1998
- The HTK book. Steve Young et al.



- Fundamentals of speech recognition. L. Rabiner. Prentice Hall. 1993.
- A.B. Downey, F. W. Olin. ThinkDSP: Digital Signal Processing in Python. O'Reilly Media, 2016. ISBN: 9781491938454 url: <https://greenteapress.com/thinkdsp/thinkdsp.pdf>
- K. Stevens, Acoustic Phonetics, MIT Press, 1998