COURSE DATA

Data Subject	
Code	36439
Name	Audio and voice analysis
Cycle	Grade
ECTS Credits	4.5
Academic year	2022 - 2023

Study (s)

Degree	Center	Acad. year	Period
1406 - Degree in Data Science	School of Engineering	4	Second term

Subject-matter

Degree	Subject-matter	Character
1406 - Degree in Data Science	16 - Audio and Speech Analysis and Processing	Optional

Coordination

Name	Department
ROMERO GOMEZ, VERONICA	240 - Computer Science
SEGURA GARCIA, JAUME	240 - Computer Science

SUMMARY

The subject Audio and Voice Analysis and Processing is a fourth year subject that is part of the optional set of subjects offered in the Degree in Data Science. The course complements the contents seen in other subjects of the Degree such as Signals and Systems, Machine Learning and Natural Language Processing, offering an applied vision of the concepts studied in the course, offering an applied vision of the concepts studied throughout these subjects. Thus, the topics covered by this subject are oriented towards the application of digital signal processing in the field of audio and voice processing. The course justifies the importance of digital audio signal processing in current multimedia and mass data processing systems, briefly reviewing some of the basic concepts studied in previous courses. The structuring of the contents follows an approach that begins by introducing the more theoretical concepts in the first part of the course, giving way to topics that use such concepts in current practical systems. Thus, we start with a review of the basic concepts of signal processing, as well as other relevant concepts in audio such as sampling, quantification, short time analysis/synthesis techniques and statistical processing methods. It continues with the study of the human auditory system and its impact on the design of practical lossy audio coding systems, the fundamentals of the phonatory system and the source/filter models used in

speech coding. Concepts such as optimal filtering, linear prediction or sub-band analysis are introduced from an applied point of view. The theoretical part is complemented by the fundamentals of speech recognition systems based on hidden Markov models and language models. In addition, machine learning techniques with application to speaker recognition and speech synthesis are studied. The aim of the course is to provide the student with an applied vision of the concepts studied throughout the degree, facilitating the understanding of abstract terms through the study of concrete applications. Thus, the aim is for the student to obtain a practical vision of the study of the processing and characterisation of the voice signal, language models, machine learning, voice recognition and synthesis.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

COMPETENCES (RD 1393/2007) // LEARNING OUTCOMES (RD 822/2021)

1406 - Degree in Data Science

- (CG01) Knowledge of basic subjects and technologies that enable students to learn new methods and technologies, and to provide them with versatility to adapt to new situations.
- (CT01) To be able to access (bibliographical) information tools and appropriately use them in the development of their daily tasks.
- (CE03) Ability to solve classification, modelling, segmentation and prediction problems from a set of data.
- (CE10) Ability to digitally process signals and extract information from them.
- (CB5) Students must have developed the learning skills needed to undertake further study with a high degree of autonomy.

LEARNING OUTCOMES (RD 1393/2007) // NO CONTENT (RD 822/2021)

To apply estimation and detection techniques to audio signals (CB05, CE10), understanding the statistical fundamentals associated with them. To know the fundamentals of speech recognition systems based on hidden Markov models. To apply machine learning techniques to audio and voice signals, with application to speaker recognition and other classification problems(CB05, CE03). To understand the basic fundamentals of the generation and perception of audio signals in general, voice and music (CB05, CE10). To know the representation of audio signals in the time and frequency domain, as well as the mathematical foundations associated with both domains. To apply time-frequency transformations for short-time analysis of time series (CB05, CE10). To apply estimation and detection techniques on audio signals, understanding the statistical fundamentals associated with them. (CE03, CE10).

DESCRIPTION OF CONTENTS

1. Introduction

Overview of digital audio, image and video signal processing. Signal processing in multimedia systems.

2. Review of signal processing

Introduction. Review of convolution and filtering. Sampling theory. Discrete-time signal processing. DFT and DTFT. Random processes

3. Audio analysis and synthesis.

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

4. Coding

Introduction. Linear prediction. Subband coding and transforms MPEG Audio

5. Speech recognition introduction.

Problem definition. Evaluation. Problems classification. History.

6. Speech recognition.

Statistical approximation. Hidden Markov Models. Language modelling. Decoding.

7. Speech recognition using neural networks

Recurrent Neural Networks, Long short-term memory, Connectionist Temporal Classification, Convolutional Neural Networks, RAH using neural networks.

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	22,00	100
Laboratory practices	15,00	100
Classroom practices	8,00	100
Attendance at events and external activities	2,00	0
Development of group work	6,00	0
Development of individual work	36,50	0
Preparation of evaluation activities	2,00	0
Preparation of practical classes and problem	6,00	0
Resolution of case studies	15,00	0
TOTAL	**112,50**	

TEACHING METHODOLOGY

MD1 - Theoretical activities. Expository development of the subject with the participation of the student in the resolution of specific questions. Individual evaluation questionnaires.
MD2 - Practical activities. Learning by solving problems, exercises and case studies through which competences on the different aspects of the subject are acquired.
MD4 - Laboratory and/or computer classroom work. Learning by means of activities carried out individually or in small groups and carried out in laboratories and/or computer classrooms.

1) Classroom work consisting of:
a) Theory classes, which will consist of the presentation and basic explanation of the corresponding subject. Short activities will be proposed, which will require the intervention of the students with the aim of confirming the understanding of the theory presented.
b) Exercise classes, designed to solve problems of greater temporal or conceptual scope.
c) Laboratory classes, designed to experimentally test some of the most relevant issues seen in the theory classes.

2) Non-attendance work consisting of:
a) Resolution and presentation of exercises. This involves solving the exercises proposed by the lecturer and/or the public presentation of the resolution of some of them.
b) Preparation of exams.
c) Preparation of the laboratory practicals, for which the student must have read and assimilated the content of the practical report, as well as having reviewed the relevant theory.

Individual and/or group tutorials. Certain hours of unscheduled tutorials per week are established, which students may attend to clarify their doubts.

EVALUATION

SE1 - Objective test, consisting of one or several exams consisting of both theoretical-practical questions and problems. 30%+30%
SE2 - Evaluation of practical activities based on the elaboration of papers/memorandums and/or oral presentations. 35%
SE3 - Continuous assessment of each student, based on the participation and degree of involvement of the student in the teaching-learning process, taking into account regular attendance at the scheduled face-to-face activities and the resolution of questions and problems proposed periodically. 5%

The evaluation mechanism is what could be called an adapted traditional one, which does not amount to a complete continuous evaluation. The following items and assessments are taken into account:
Assessment of attendance and participation (5% of the final mark). The mark obtained in the second sitting is retained.
Two partial exams (30% of the final mark each one).
Attendance and completion of the practicals (20% of the final mark). Grade not recoverable, the grade obtained at the second sitting is retained.
Resolution of proposed exercises (15% of the final mark). Grade not recoverable, the grade obtained in the second round is retained.

A student is considered to be attending class regularly when he/she has not missed more than 25% of the classes given up to half of the time the subject is taught.
For students who are unable to attend class regularly, an alternative model is offered in which the assessment of attendance and participation will be replaced by additional work with an equivalent total percentage. Students who wish to make use of this option must inform the teacher during the first 3 weeks of the course.

The minimum required to pass the course will be the equivalent of a 3.5 out of 10 in both partial exams. The rest of the evaluable items are not subject to a minimum.
In both first and second call, those students who do not pass the 5 by means of the continuous assessment activities will be able to choose between two options:
- Final exam (60%). The mark corresponding to the partial exams will be recovered with a single final exam.
- Partial exam (30%/30%). The mark of the partial exam higher than 3.5 is kept and the failed part is recovered by means of an exam.

REFERENCES

Basic

- 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.

Additional

- A.B. Downey, F. W. Olin. ThinkDSP: Digital SignalProcessingin Python. O'ReillyMedia, 2016. ISBN: 9781491938454 url: https://greenteapress.com/thinkdsp/thinkdsp.pdf
- K. Stevens, Acoustic Phonetics, MIT Press, 1998