Course guide
320123 - TP - Speech Technology

Last modified: 19/04/2023

Unit in charge: Terrassa School of Industrial, Aerospace and Audiovisual Engineering
Teaching unit: 739 - TSC - Department of Signal Theory and Communications.

Degree: BACHELOR'S DEGREE IN AUDIOVISUAL SYSTEMS ENGINEERING (Syllabus 2009). (Optional subject).

Academic year: 2023 **ECTS Credits:** 6.0 **Languages:** Catalan, Spanish

LECTURER

Coordinating lecturer: Ignasi Esquerra

Others: Albino Nogueiras

PRIOR SKILLS

It is recommended to have taken the courses in "Digital Audio Processing" and "Algorithms and Audiovisual Programming", and to have a certain knowledge in programming in Linux platforms.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CE26. (ENG) AUD: Coneixements i capacitats per aprofundir en tecnologies específiques de l¿àmbit.

TEACHING METHODOLOGY

Learning with a teacher:
a) Laboratory sessions: Theory explanations and practical demostrations. Hands-on exercises. Evaluation.
Autonomous learning:
b) Individual study, exercise solving and handouts preparation.

LEARNING OBJECTIVES OF THE SUBJECT

Understanding the technologies used in human-computer interaction systems by voice. Students by the end of the course will learn several techniques of speech signal processing and will be able to put them into practice in speech recognition and speech synthesis systems.

STUDY LOAD

Type	Hours	Percentage
Hours large group	30,0	20.00
Hours small group	30,0	20.00
Self study	90,0	60.00

Total learning time: 150 h

CONTENTS

1. HUMAN-COMPUTER INTERACTION BY VOICE

Description:
Spoken language technologies. Historical review of Human-Computer interaction by voice. Applications and current research lines.

Full-or-part-time: 10h
Theory classes: 2h
Laboratory classes: 2h
Self study : 6h

2. DATABASES

Description:
Text and speech databases. Phonetics and linguistics. Production and perception of speech. Representation and analysis of speech signals.

Full-or-part-time: 30h
Theory classes: 8h
Laboratory classes: 8h
Self study : 14h

3. SPEECH SYNTHESIS

Description:
Text analysis. Phonetic transcription. Prosody. Unit-selection synthesis. Probabilistic model synthesis.

Full-or-part-time: 46h
Theory classes: 8h
Laboratory classes: 8h
Self study : 30h

4. SPEECH RECOGNITION

Description:
Parametrization. Acoustic models. Hidden Markov Models. Language models.

Full-or-part-time: 64h
Theory classes: 12h
Laboratory classes: 12h
Self study : 40h

GRADING SYSTEM

The final mark is the weighted sum of partial marks of oourse units. Each unit is assessed with several laboratory assignments or exams. None of the qualification elements has a weight over 25%.
Unit 1 (10%), Unit 2 (20%), Unit 3 (35%), Unit 4 (35%)

BIBLIOGRAPHY

Complementary:
- O'Shaughnessy, D. Speech communications: human and machine. 2nd ed. New York: IEEE Press, 2000. ISBN 978-0780334496.
- 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.