Speech and Audio Processing

Academic Year: (2019 / 2020) **Review date: 19-03-2019**

Department assigned to the subject:

Coordinating teacher: GALLARDO ANTOLIN, ASCENSION

Type: Electives ECTS Credits : 6.0

Year : 1 Semester : 1

OBJECTIVES

- Knowledge on the speech production mechanism and the linguistic categories of the voice
- Knowledge on sound perception
- Sound knowledge on implementation and fundamentals of speech and audio coders, speech recognition, speech synthesis, speaker recognition and audio classification
- Sound knowledge on coding standards and metadata
- Knowledge on VoIP, dialog systems, voice-guided applications and computer-telephony integrated systems
- Capability to initiate research work in the following fields: speech and audio coding, speech recognition, speech synthesis, speaker recognition and audio classification

DESCRIPTION OF CONTENTS: PROGRAMME

Unit 0. Introduction to Speech Technologies
Unit 1. The Auditory System and Speech Perception
Unit 2. The Speech Production System and Phonation. Speech and Audio Coding
Unit 3. Automatic Speech Recognition
Unit 4. Fundamentals of Speech Enhancement
Unit 5. Speaker Recognition
Unit 6. Applications

LEARNING ACTIVITIES AND METHODOLOGY

The following learning activities and methodologies are combined:

- Theory classes
- Guided lab assignments
- Research papers' presentations
- Final project

ASSESSMENT SYSTEM

First call:
- Research paper presentation (30%)
- Comprehension tests (30%)
- Final project (40%)

Second call:
- Research paper presentation (30%)
- Comprehension tests (30%)
- Final project (40%)

% end-of-term-examination:	0
% of continuous assessment (assigments, laboratory, practicals…):	100

BASIC BIBLIOGRAPHY

- B. Gold and N. Morgan, Speech and Audio Signal Processing: Processing and Perception of Speech and Music,, New York, John Wiley & Sons,, 2000

- D. O'Shaughnessy, Automatic speech recognition: History, methods and challenges,, Pattern Recognition, 41 (10) pp. 2965-2979, , 2008
- D. O'Shaughnessy, Speech Communication: Human and Machine (Second Edition),, New York: IEEE Press,, 2000

- K.C. Pohlmann, Principles of Digital Audio (Fifth Edition),, New York: MCGraw-Hill,, 2005

- S. Huang, A. Acero, H.W. Hon, Spoken Language Processing: A Guide to Theory, Algorithms and System Development,, New Jersey: Prentice Hall,, 2001

ADDITIONAL BIBLIOGRAPHY

- F. Charpentier and E. Moulines, Pitch-synchronous Waveform Processing Techniques for Text-to-speech Synthesis Using Diphones,, Proc. of the First European Conference on Speech Communication and Technology (EUROSPEECH¿89), pp. 2013-2019,, 1989
- H. Hermansky, Should Recognizers Have Ears?,, In Proceedings of ESCA Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels, pp.1-10, France,, 1997
- H. Misra, J. Vepa and H. Bourlard, Multi-stream ASR: Oracle Test and Embedded Training,, IDIAP Technical Report, IDIAP-RR 05-62,, 2005
- Hermansky, H.; Morgan, N.; Bayya, A.; Kohn, P., RASTA-PLP speech analysis technique,, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol.1, no., pp.121-124 vol.1, 23-26,, 1992

- Ian Vince McLoughlin, Line Spectral Pairs,, Signal Processing 88, pp. 448-467,, 2008

- Karlheinz Brandenburg, MP3 AND AAC Explained,, AES 17th International Conference on High Quality Audio Coding,, 1999
- M. J. F. Gales and S. J. Young, Robust Continuous Speech Recognition Using Parallel Model Combination,, IEEE Transactions on Speech and Audio Processing, vol. 4, no. 5, pp. 352-359,, September 1996

- M.F. Schroeder and B.S. Atal, Code-Excited Linear Prediction (CELP) high-Quality Speech at Very Low Bit Rates,, ICASSP-1985, pp. 937-940, , 1985
- N. Morgan , H. Bourlard, Continuous Speech Recognition using Multi-Layer Perceptrons with Hidden Markov Models,, Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, Albuquerque, vol. 1, pp. 413-416, , 1990
- P. Warnestal, Modeling a Dialogue Strategy for Personalized Movie Recommendations,, Proceedings of the Beyond Personalization 2005 workshop on the Next Stage of Recommender Systems Research, pages 77-82, , 2005

- Qifeng Zhu, Barry Chen, Nelson Morgan and Andreas Stolcke, Tandem Connectionist Feature Extraction for Conversational Speech Recognition,, In "Machine Learning for Multimodal Interaction", pp- 223-231, , 2005

- Reynolds, D.A.; Rose, R.C., Robust Text-independent Speaker Identification Using Gaussian Mixture Speaker Models,, IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 72-83,, January 1995

- T. Hazen, T. Burianek, J. Polifroni and S. Seneff, Recognition Confidence Scoring for Use in Speech Understanding Systems,, Proc. ISCA Tutorial and Research Workshop ASR2000,, September 2000