

Curso Académico: (2019 / 2020)

Fecha de revisión: 19-03-2019

Departamento asignado a la asignatura:

Coordinador/a: GALLARDO ANTOLIN, ASCENSION

Tipo: Optativa Créditos ECTS : 6.0

Curso : 1 Cuatrimestre : 1

OBJETIVOS

- Conocimiento de los mecanismos de producción de habla y las categorías lingüísticas que describen el habla
- Conocimiento sobre percepción del sonido
- Conocimiento operativo acerca de la implementación y funcionamiento de sistemas de codificación de voz y audio, reconocimiento de habla, conversión texto-voz, reconocimiento de locutores y clasificación de audio
- Conocimientos operativos acerca de estándares de codificación y metadatos
- Conocimiento acerca del funcionamiento de aplicaciones de VoIP, sistemas de diálogo, aplicaciones guiadas por voz y sistemas integrados de voz telefónica y ordenadores
- Capacidad para iniciar tareas de investigación en los ámbitos de codificación de voz o audio, reconocimiento de habla, conversión texto-habla, reconocimiento de locutores y clasificación de audio

DESCRIPCIÓN DE CONTENIDOS: PROGRAMA

- Tema 0. Introducción a las tecnologías del habla
- Tema 1. El sistema auditivo y percepción del habla
- Tema 2. El sistema de producción de voz y fonación. Codificación de voz y de audio
- Tema 3. Reconocimiento automático de habla
- Tema 4. Fundamentos de mejora de voz
- Tema 5. Reconocimiento de locutor
- Tema 6. Aplicaciones

ACTIVIDADES FORMATIVAS, METODOLOGÍA A UTILIZAR Y RÉGIMEN DE TUTORÍAS

La asignatura se desarrolla con una combinación de las siguientes actividades formativas y metodologías:

- Clases magistrales
- Prácticas guiadas de laboratorio
- Presentación de artículos de investigación
- Proyecto final

SISTEMA DE EVALUACIÓN

Convocatoria ordinaria:

- Presentación de artículos de investigación (30%)
- Test de comprensión de artículos de investigación (30%)
- Proyecto final (40%)

Convocatoria extraordinaria:

- Presentación de artículos de investigación (30%)
- Test de comprensión de artículos de investigación (30%)
- Proyecto final (40%)

Peso porcentual del Examen Final:

0

Peso porcentual del resto de la evaluación:

100

BIBLIOGRAFÍA BÁSICA

- 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

BIBLIOGRAFÍA COMPLEMENTARIA

- 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