



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

GUÍA DOCENTE: Tecnologías del Habla (ID)

Curso Académico: 2017-2018

Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Centro: Escuela Politécnica Superior
Universidad: Universidad Autónoma de Madrid

Última modificación: 08/07/2017
Estado: Pendiente de aprobación



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

1. ASIGNATURA (ID)

Tecnologías del Habla (ID)

1.1. Programa

Máster Universitario en Ingeniería de Telecomunicación (ing.tel)

1.2. Código asignatura

32651

1.3. Área de la asignatura

TSC

1.4. Tipo de asignatura

Optativa

1.5. Semestre

Segundo semestre

1.6. Créditos

3 ETCS

1.7. Idioma de impartición

Las transparencias se proporcionarán en castellano y en inglés, las clases se impartirán en castellano, y el material de estudio (capítulos de libro y artículos de revista) estará únicamente en inglés.

1.8. Recomendaciones / Requisitos previos

Para esta asignatura, es imprescindible estar cursando en la asignatura “Procesado Avanzado de Señales Multimedia” de segundo semestre de este Máster. También es recomendable haber cursado la asignatura de primer semestre “Teoría de la Información para Comunicaciones”. Asimismo, es recomendable tener conocimientos



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

previos en procesamiento de señales de audio y voz, al nivel exigido para la admisión al presente Máster.

1.9. Datos del equipo docente

Coordinador:

Dr. Joaquín González Rodríguez (Coordinador)
Departamento de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior. Despacho: C-204
Tel.: +34 914973142
e-mail: joaquin.gonzalez@uam.es
Web: <http://atvs.ii.uam.es/listpeople.do#Gonzalez-Rodriguez>

1.10. Objetivos del curso

En esta asignatura se estudian los principales métodos, técnicas y sistemas dentro del tratamiento digital de audio y voz y las tecnologías del habla orientados a reconocer personas por sus voces. El curso considera tanto los métodos básicos comunes a todas las aproximaciones tecnológicas descritas, como las técnicas particulares más utilizadas en cada una de dichas aplicaciones. Las sesiones prácticas del curso están orientadas en la implementación de etapas presentes en estos sistemas, como la extracción de características para reconocimiento de locutores o el modelado de dichas características.

La competencia específica que se espera que los estudiantes adquieran con esta asignatura es la siguiente:

PR1 - Capacidad para la integración de tecnologías y sistemas propios de la Ingeniería de Telecomunicación, con carácter generalista, y en contextos más amplios y multidisciplinares como por ejemplo en bioingeniería, conversión fotovoltaica, nanotecnología, telemedicina.

A continuación, se especifican los objetivos generales y específicos de la asignatura que pretenden alcanzarse.

OBJETIVOS GENERALES	
G1	Entender los distintos tipos de información contenidos en la señal de voz y los procedimientos eficientes de extracción de dicha información
G2	Entender las distintas formas de modelar y comparar secuencias de patrones temporales
G3	Entender las tecnologías y arquitecturas necesarias para realizar reconocimiento del mensaje tanto en habla aislada como habla continua
G4	Entender las tecnologías y arquitecturas necesarias para realizar reconocimiento de locutor tanto dependiente como independiente de texto
G5	Entender las tecnologías y arquitecturas necesarias para realizar reconocimiento del idioma hablado basado tanto en sistemas



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

OBJETIVOS GENERALES	
	fonético/lingüísticos como acústico/espectrales

OBJETIVOS ESPECÍFICOS	
TEMA 1.- Análisis localizado y parametrización de señales de voz	
1.1.	Entender la necesidad y fundamentos del análisis espectral localizado a corto plazo
1.2.	Manejar los distintos esquemas de parametrización más representativos (LPC, MFCC, pitch, energía, cruces por cero)
TEMA 2.- Modelado y comparación de secuencias temporales	
2.1.	Entender la necesidad y mecanismos para alinear diferentes secuencias temporales
2.2.	Manejar las técnicas básicas de modelado (DTW, VQ, GMM)
TEMA 3.- Sistemas de reconocimiento de voz	
3.1.	Entender la necesidad de manejar estados ocultos para representar secuencias de voz mediante modelos ocultos de Markov (HMM)
3.2.	Manejar modelos acústicos y de lenguaje, así como algoritmos y herramientas como token-passing y lattices
TEMA 4.- Sistemas de reconocimiento de locutor	
4.1.	Capacidad de implementar sistemas dependientes e independientes de texto con distintas arquitecturas
4.2.	Conocer esquemas de compensación de variabilidad inter-sesión basados en sub-espacios, como JFA, i-Vectors o PLDA.
TEMA 5.- Sistemas de reconocimiento de idioma	
5.1.	Distinguir sistemas fonético/lingüísticos y acústico/espectrales
5.2.	Conocer esquemas de reconocimiento de idioma basados en redes neuronales profundas



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

Contenidos del programa

TEORÍA

1. Análisis localizado y parametrización de señales de voz *(semana 1)*
2. Modelado y comparación de secuencias temporales: *(semanas 2 y 3)*
 - a. Alineamiento temporal dinámico (DTW)
 - b. Cuantificación vectorial (VQ)
 - c. Modelos de mezclas de gaussianas (GMM)
3. Sistemas de reconocimiento de voz *(semanas 3 y 4)*
 - a. Modelos Ocultos de Markov (HMM)
 - b. Modelos acústicos y de lenguaje
 - c. Reconocimiento de habla continua: token-passing y celosias (lattices)
4. Sistemas de reconocimiento de locutor *(semanas 5 y 6)*
 - a. Evaluación de sistemas de detección: ROC, DET, costs de error
 - b. Sistemas dependientes de texto
 - i. Clave fija versus text-prompted
 - ii. Modelos de palabra/frase versus concatenación dinámica de unidades
 - c. Sistemas independientes de texto
 - i. Extracción de información de bajo y alto nivel
 - ii. VQ, GMM, kernel-SVM, SuperVector-SVM y combinación de clasificadores
 - iii. i-vectors y compensación de variabilidad intersesión
5. Sistemas de reconocimiento de idioma *(semana 7)*
 - a. Sistemas fonético/lingüísticos
 - i. Modelado estadístico de secuencias de fonemas independientes de idioma (PPRLM)
 - ii. Reconocedores de voz dependientes de idioma en paralelo (PPR)
 - b. Sistemas acústico/espectrales
 - i. GMM, kernel-SVM, SuperVector-SVM
 - ii. i-vectors y compensación de variabilidad intersession
 - c. Redes Neuronales Profundas (DNN)
 - i. Shallow and Deep Feed-Forward Neural Networks
 - ii. Redes Neuronales Recurrentes: LSTMs

LABORATORIO

- | | | |
|----|---|-------------------|
| P1 | Análisis localizado y parametrización de señales de voz | <i>(semana 1)</i> |
| P2 | Programación de VQ (k-means y/ó binary splitting) | <i>(semana 2)</i> |
| P3 | Programación de DTW y reconocimiento de dígitos | <i>(semana 3)</i> |
| P4 | Construcción y evaluación de HMMs con HTK | <i>(semana 4)</i> |
| P5 | Reconocimiento de locutor con GMM | <i>(semana 5)</i> |
| P6 | Supervectores, subespacios e i-vectors | <i>(semana 6)</i> |
| P7 | Scoring avanzado y evaluación de sistemas | <i>(semana 7)</i> |



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

1.11. Bibliografía

Bibliografía básica:

1. Benesty, Jacob; Sondhi, M. M.; Huang, Yiteng (Eds.). Springer Handbook of Speech Processing. 2008.
 - a. Tema 1 (parametrización): capítulos 2, 7, 9, 10 y 11
 - b. Tema 3 (reconocimiento de voz): capítulos 27, 31, 32 y 35
 - c. Tema 4 (locutor): capítulos 36, 37 y 38
 - d. Tema 5 (idioma): capítulos 39 y 41

Bibliografía complementaria:

2. Xuedong Huang, Alex Acero, and Hsiao-Wuen Hon, Spoken Language Processing, pp. 1008, Prentice-Hall, May 2001.
3. Discrete-Time Speech Signal Processing. Principles and Practice Thomas F. Quatieri. Prentice-Hall Ed. 2002.
4. Digital Processing of Speech Signals L.R. Rabiner, R.W. Schafer. Prentice-Hall Ed. 1978.
5. Pattern Classification R.O. Duda, P.E. Hart, D.G. Stork, 2nd Edition, John Wiley and Sons, NY, 2001.
6. C. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.

Transparencias:

Se facilitarán transparencias de cada tema. No obstante, las transparencias son sólo el guión de clase y los contenidos detallados se encuentran en los capítulos correspondientes del libro de referencia.

1.12. Metodología docente

Las clases presenciales programadas en esta asignatura están orientadas a las explicaciones teóricas, a las prácticas en laboratorio y a la evaluación continua. La actividad presencial se divide, de acuerdo con el horario de la asignatura, en cuatro horas semanales, dos de teoría y dos de laboratorio.

Actividad en el aula:

La actividad en el aula se encuentra repartida en tres aspectos: explicaciones teóricas, casos prácticos y evaluación continua.

La evaluación continua consistirá en la realización de dos pruebas de una hora cada una, a realizar en el horario de teoría de la asignatura. La primera de ellas en la semana 4, y la segunda en la semana 7.



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

Las explicaciones teóricas resumirán los conceptos más importantes de cada tema, pero se considera esencial que los estudiantes profundicen posteriormente empleando la bibliografía de la asignatura. Esto último es especialmente importante dado el reducido número de horas presenciales con respecto a los ECTS con los que la asignatura cuenta.

Actividad en el laboratorio:

Las sesiones de prácticas son de asistencia obligatoria, y están diseñadas para realizarse en su mayor parte en el horario de laboratorio. Sin embargo, se cuenta con que las prácticas se pueden completar fuera del horario de laboratorio, de acuerdo con la carga programada para ello, tanto para mejoras o extensiones de las mismas, como para la preparación de las entregas (memorias y/o resultados). Las sesiones prácticas se desarrollarán, a poder ser, de forma individual, aunque en caso de existir limitaciones logísticas se formarán parejas. En cualquier caso, la evaluación es completamente individual.

Las prácticas consistirán en desarrollos prácticos para reforzar lo aprendido en las sesiones de teoría.

1.13. Trabajo del estudiante

		Horas	%	Horas	%
Presencial	Clases - Explicación teoría	8	10,7%	28	37,3%
	Clases - Casos prácticos	4	5,3%		
	Clases - Prácticas	14	18,7%		
	Pruebas de evaluación (*)	2	2,7%		
No presencial	Estudio semanal regulado	14	18,7%	47	62,7%
	Finalización de prácticas y/o memorias	21	28%		
	Preparación exámenes (*)	12	16%		
Carga total de horas de trabajo: 25 horas x 3 ECTS		75	100%	75	100%

(*) Incluye pruebas de evaluación continua y finales de teoría.

1.14. Métodos de evaluación y porcentaje en la calificación final

La evaluación de la asignatura, o nota final (NF), dependerá de la nota de teoría (TE) y de la nota de prácticas (PR), en la siguiente proporción:

$$NF = 0.3*TE + 0.7*PR$$



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

Ambas partes, TE y PR se puntúan sobre 10 puntos. Es necesario obtener una calificación mínima de 5 puntos en TE y PR para poder aplicar la expresión anterior. Si no se cumple esta condición, la calificación numérica que se hará constar en actas será:

$$NF = 0.3 * \text{Mín}(5, TE) + 0.7 * \text{Mín}(5, PR)$$

Si un estudiante recibe la calificación de “no evaluado” en teoría o en prácticas, su nota final será “no evaluado”.

Nota de teoría, TE:

La nota de teoría será el resultado de uno de los dos procesos de evaluación que se describen:

- 1- Evaluación continua (TE-C): la realización de 2 pruebas de evaluación continua (EC1 y EC2, puntuadas cada una sobre 10 puntos) planificadas en el transcurso de la asignatura (ver más adelante).
- 2- Evaluación única (TE-U): la realización de una prueba o examen final (EF) planificado en la convocatoria ordinaria o en la convocatoria extraordinaria de la asignatura.

Todas las pruebas de evaluación se realizarán sin libros ni apuntes, y no se permitirá el uso de teléfonos móviles, etc. Cada prueba de evaluación consistirá en un conjunto de preguntas y/o ejercicios que cubran de forma uniforme el temario impartido.

La *evaluación continua* será el proceso asumido por defecto. El resultado de este proceso será una media de las pruebas realizadas, de acuerdo con la siguiente fórmula:

$$TE=TE-C= 0.5*EC1 + 0.5*EC2$$

La nota de todas las pruebas EC1 y EC2 debe ser mayor o igual que 3.0. En caso de que al menos una de las pruebas tenga una calificación inferior a 3.0, se aplicará la siguiente fórmula:

$$TE=TE-C= 0.5*\text{Mín}(5, EC1) + 0.5*\text{Mín}(5, EC2)$$

El temario que se evalúa en cada prueba de evaluación continua es el siguiente:

- EC1: Temas 1, 2 y 3.
- EC2: Temas 4 y 5.

La *evaluación única* es el proceso excepcional dirigido a estudiantes que o bien no siguen el proceso de evaluación continua (TE-C=0), o bien, habiéndolo seguido, optan



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

por presentarse a un examen final para aprobar o aumentar su nota. En este caso la calificación se obtendrá según:

$$TE=TE-U= \text{Max}(EF, TE-C)$$

La calificación de teoría sólo se conserva para la convocatoria extraordinaria del mismo curso académico.

Los estudiantes que no realicen un examen final (EF) ni al menos 1 de las 2 pruebas de evaluación continua recibirán la calificación de “no evaluado”.

Nota de prácticas, PR:

La nota de prácticas será el resultado de uno de los dos procesos de evaluación que se describen:

- 1- Evaluación continua (PR-C): valoración promedio de los resultados obtenidos en cada una de las prácticas.
- 2- Evaluación única (PR-U): la realización de una prueba o examen final de prácticas (PRF) planificado en la convocatoria ordinaria o en la convocatoria extraordinaria de la asignatura.

La **evaluación continua** será el proceso asumido por defecto. Sólo se podrá evaluar el laboratorio mediante evaluación única en circunstancias excepcionales debidamente justificadas, y tras acuerdo previo con el profesor.

La asistencia al laboratorio es obligatoria. Cada práctica será evaluada mediante entregas de resultados y/o memorias, dependiendo de la práctica en cuestión. De esta forma, PR-C será el promedio de todas las notas de cada una de las prácticas.

La **evaluación única** es el proceso excepcional dirigido a estudiantes que por causas excepcionales debidamente justificadas no siguen el proceso de evaluación continua (PR-C=0). También es la única posibilidad de evaluación en convocatorias extraordinarias.

Nota importante sobre los exámenes prácticos de tipo PR-U, tanto en convocatoria ordinaria como extraordinaria: por motivos de organización, ha de solicitarse expresamente al coordinador de la asignatura, incluyendo la documentación justificativa pertinente, al menos dos semanas antes del inicio de los exámenes de la convocatoria correspondiente.

El examen final de prácticas (PRF) consistirá en la realización, durante 3 horas, de varios ejercicios prácticos de programación en Matlab cubriendo la diversidad de contenidos del temario. No se permitirá el uso de código desarrollado con anterioridad, libros ni apuntes.

La calificación final de prácticas se obtendrá según:



Asignatura: Tecnologías del Habla (ID)
Código: 32651
Institución: Escuela Politécnica Superior
Programa: Máster Universitario en Ingeniería de Telecomunicación (ing.tel)
Nivel: Máster
Tipo: Optativa
ECTS: 3

$$PR = \text{Max}(PRF, PR-C)$$

Si por motivos de asistencia (ver arriba) un estudiante es declarado NO APTO en prácticas, su nota de prácticas será 0.

La calificación de prácticas sólo se conserva para la convocatoria extraordinaria del mismo curso académico.

ATENCIÓN: Cualquier copia descubierta que se haya realizado a lo largo del curso, tanto en cualquiera de las actividades de teoría desarrolladas, como en cualquiera de los apartados de las prácticas, serán penalizadas con rigurosidad, siguiendo las normativas de la UAM y de la EPS.

1.15. Planificación / Cronograma

El siguiente cronograma indica la distribución *orientativa* de contenido, incluyendo la programación de las pruebas de evaluación sobre dicho contenido.

- Semana 1:
 - Teoría: tema 1 (parametrización)
 - Laboratorio: práctica 1 (parametrización)
- Semana 2:
 - Teoría: tema 2 (modelado)
 - Laboratorio: práctica 2 (VQ)
- Semana 3:
 - Teoría: tema 2 (modelado) y tema 3 (reconocimiento de voz)
 - Evaluación: práctica 3 (DTW)
- Semana 4:
 - Teoría: tema 3 (reconocimiento de voz) y 1er. examen parcial
 - Laboratorio: práctica 4 (HMM)
- Semana 5:
 - Teoría: tema 4 (locutor)
 - Evaluación: práctica 5 (GMM)
- Semana 6:
 - Teoría: tema 4 (locutor)
 - Evaluación: práctica 6 (i-vectors)
- Semana 7:
 - Teoría: tema 5 (idioma) y 2o. examen parcial
 - Evaluación: práctica 7 (scoring y evaluación de sistemas)