



Pla Docent de l'Assignatura

Guia Docent

Nom de l'assignatura: Processament de la Parla

Curs acadèmic: 2011-2012

Curs: Segon **Trimestre:** Tercer

Estudis: Grau en Enginyeria en Informàtica (Optativa), Grau en Enginyeria en Telemàtica (Optativa) i Grau en Enginyeria en Sistemes Audiovisuals (Obligatòria)

Codi assignatura: 21480-21762-21610

Nombre de crèdits ECTS: 4

Nombre total d'hores de dedicació: 100 hores

Llengua o llengües de docència: Català (explicacions), anglès (material)

Professorat: Emilia Gómez, Waldo Nogueira, Martí Umbert

Professorat responsable: Emilia Gómez

1. Dades descriptives de l'assignatura

Nom de l'assignatura:Processament de la Parla

Curs acadèmic:2011-2012

Curs:Segon **Trimestre:** Tercer

Estudis: Grau en Enginyeria en Informàtica (Optativa), Grau en Enginyeria en Telemàtica (Optativa) i Grau en Enginyeria en Sistemes Audiovisuals (Obligatòria)

Codi assignatura:21480-21762-21610

Nombre de crèdits ECTS: 4

Nombre total d'hores de dedicació:100 hores

Llengua o llengües de docència: Català (explicacions), anglès

(material)Professorat:Emilia Gómez, Waldo Nogueira, Martí Umbert

Professorat responsable:Emilia Gómez, Waldo Nogueira, Martí Umbert

2. Presentació de l'assignatura

Aquest és un curs intermitj al processament digital del senyals sonores, pensat per estudiants de Enginyeria en Sistemes Audiovisuals.

El curs està construït sobre la base de assignatures prèvies de la titulació, principalment Enginyeria Acústica i Senyals i Sistemes (segón curs, Enginyeria en Sistemes Audiovisuals).

Ens centrarem en l'estudi de les principals tècniques d'anàlisi, descripció, síntesi i processament de senyals de veu.

3. Competències a assolir a l'assignatura

Competències¹ a treballar a l'assignatura segons l'indicat en el pla d'estudis del grau.

Competències transversals	Competències específiques
<i>Instrumentals</i> G1. Capacitat d'anàlisi i síntesi G2. Capacitat d'organització i planificació G3. Capacitat per aplicar els coneixements a l'anàlisi de situacions i la resolució de problemes G4. Habilitat en la cerca i la gestió de la informació G5. Habilitat en la presa de decisions Capacitat de comunicar-se en context acadèmic de forma oral i escrita en català, castellà i anglès	<i>Competències Específiques de Formació Bàsica</i> B4-INF. Capacitat per analitzar funcions de variable complexa. B7-INF. Conèixer les transformades de Fourier per a senyals analògics i digitals i els elements bàsics de la teoria del mostreig de senyals. B8-INF. Capacitat per resoldre sistemes lineals i invariants i les funcions i transformades relacionades. B7-T. Comprendre i utilitzar els principis de la probabilitat, els B4-A. Capacitat per analitzar funcions de variable complexa. B7-A. Conèixer les transformades de Fourier per a senyals analògics i digitals. Conèixer els elements bàsics de la

¹**Competències transversals:** Aquelles que es requereixen en l'exercici de qualsevol titulació o carrera (comunicació verbal i escrita, pensament analític i sistèmic, resolució de problemes, creativitat, etc.). Es classifiquen en:

- **Instrumentals:** Inclouen habilitats cognitives, metodològiques, tecnològiques i lingüístiques. (Ex: capacitat d'organitzar i planificar, capacitat de comunicar-se amb propietat de forma oral i escrita en català, castellà i/o anglès, tant davant d'audiències expertes com inexpertes).
- **Interpersonals:** Es defineixen com habilitats que tendeixen a facilitar els processos d'interacció social i cooperació. (Ex: capacitat per a treballar en grup, expressió del compromís ètic/social).
- **Sistèmiques o integradores:** Suposen una combinació de comprensió, sensibilitat i coneixement que permeten veure com s'agrupen i s'estableixen relacions entre les parts d'un tot. Aquestes competències requereixen, com a base, l'adquisició prèvia de competències instrumentals i interpersonals. (Ex: capacitat d'adaptar-se a nous contextos d'aprenentatge)

Competències específiques: Es relacionen amb els coneixements i pràctiques concretes del grau. (Ex: capacitat de descriure, programar, validar i optimitzar protocols i interfaces de comunicació en els diferents nivell d'una arquitectura de xarxes)

<i>Interpersonals</i> G8. Capacitat de treball en equip Capacitat de treball en contextos internacionals i interdisciplinaris. <i>Sistèmiques</i> G11. Capacitat d'aplicar amb flexibilitat i creativitat els coneixements adquirits i d'adaptar-los a contextos i situacions noves G12. Capacitat per progressar en els processos de formació i aprenentatge de manera autònoma i contínua	teoria del mostreig de senyals. Capacitat per resoldre sistemes lineals i invariants i les funcions i transformades relacionades. B9-A. Adquirir els coneixements bàsics de la física de la propagació del so i la seva relació amb els mètodes de processament de l'assenyals àudio. <i>Competències de tecnologia específica: Sistemes Audiovisuals</i> AU1. Capacitat de construir, explotar i gestionar serveis i aplicacions de telecomunicacions, enteses aquestes com a sistemes de captació, tractament analògic i digital, codificació, transport, representació, processament, emmagatzematge, reproducció, gestió i presentació de serveis audiovisuals i informació multimèdia. AU3. Capacitat de seleccionar un sistema, component o procés de l'àmbit del so i imatge per complir les especificacions requerides. AU4. Adquirir el coneixement de les tècniques de processament d'àudio necessàries per al desenvolupament de sistemes de processament acústic. AU5. Coneixement de les tècniques de tractament del senyal utilitzades en l'anàlisi, síntesi, codificació i reconeixement del parla. Desenvolupament d'aplicacions informàtiques per al tractament del parla. AU6. Coneixements de les tècniques de tractament de senyals d'àudio i música. Desenvolupament d'aplicacions informàtiques basades en el processament de música. AU22. Entendre i conèixer els principis matemàtics en els quals es basen els sistemes de codificació d'àudio i música. Entendre i conèixer els
--	---

	avantatges i inconvenients de cadascun d'aquests sistemes de codificació d'àudio i música.
--	--

4. Continguts

En aquesta assignatura es vol que l'alumne desenvolupi una metodologia per a analitzar, codificar, reconeixer i sintetitzar senyals de veu (principalment de parla) utilitzant tècniques de processament del senyal. Més concretament es pretén que l'estudiant adquireixi competències teòriques i pràctiques relacionades amb:

- Fonaments acústics, fisiològics i perceptuals de la veu.
- Fonaments de l'anàlisi digital de senyals de veu.
- Mètodes de modelat i processament de senyals de veu.
- Us de programari de processament de veu.
- Implementació, amb un llenguatge de programació, d'algorismes de processament del senyal pel processament de la veu.

Aquests conceptes s'estructuren en els següents **Blocs de Contingut**:

Bloc 1. Introducció:

- Cadena de generació/percepció de la parla.
- Aplicacions del processament digital de la parla.

Bloc 2. Fonaments:

- Fonaments acústics: mecanismes de producció de la veu, parla vs cant, classificació dels sons de la parla, representació fonètica.
- Fonaments perceptuals: alçada, intensitat i timbre.

Bloc 3. Anàlisi espectral de senyals de veu.

- Anàlisi espectral de senyals de veu: STFT, espectrograma.
- Extracció de paràmetres: energia, ZCR, ST-ACF, pitch.
- Anàlisi multi-resolució: Wavelet.

Bloc 4. Models basats en la percepció del so.

- Síntesi de veu amb modelat espectral.
- Transformacions.

Bloc 5. Models basats en la producció de la parla.

- Anàlisi Linear Predictiu (LPC).

Bloc 6. Codificació de la parla.

Bloc 7. Síntesi Text-Parla (*Text-to-Speech Synthesis*).

Bloc 8. Reconeixement de parla (*Automatic Speech Recognition*).

- Anàlisi cepstrum.
- Models de Markov.

5. Avaluació del nivell d'assoliment de les competències

Criteris generals d'avaluació

L'avaluació es fa per a cada una de les tres activitats que constitueixen l'assignatura: classes de teoria, seminaris i laboratoris. Essent:

- T: l'avaluació de la teoria mitjançant un examen final i un control de coneixements al mig del trimestre (que val un 30% de la nota de teoria). L'examen final avalua tota la matèria de l'assignatura.
- L: l'avaluació dels laboratoris amb la participació en aquests i l'entrega de les memòries de pràctiques.
- S: l'avaluació dels seminaris amb la participació en aquests.

La nota final s'obté fent la mitjana ponderada de la següent manera:

$$\text{Nota Final} = 0,5 * T + 0,4 * L + 0,1 * S$$

Al final de l'assignatura es realitza un examen escrit i individual per avaluar la comprensió dels continguts presentats en les classes de teoria i reforçats amb els seminaris i laboratoris. Aquesta avaluació és obligatòria i ha de ser qualificada amb com a mínim un 4/10 per tal de aprovar l'assignatura.

En els laboratoris es realitzen un sèrie de pràctiques que posen a prova la capacitat dels alumnes de resoldre problemes pràctics i d'implementar algorismes en forma de programes en un ordinador. Les practiques s'han d'entregar individualment a les dades programades. L'avaluació es fa a partir del seguiment realitzat a classe pel professor i de la correcció de les memòries de pràctiques. Aquesta avaluació també és obligatòria i ha de ser qualificada com a mínim amb un 5/10 per tal de aprovar l'assignatura.

Abans de cada seminari s'entrega una sèrie d'activitats als alumnes, com a una preparació prèvia al seminari. Aquestes activitats corresponen a conceptes o coneixements tractats a classe de teoria i posats a la practica en els laboratoris. Durant el seminari tots els estudiants han de participar en la resolució de les activitats. L'avaluació de seminaris es centra en la participació dels estudiants a les activitats del seminaris.

6. Bibliografia i recursos didàctics

6.1. Fonts d'informació per a l'aprenentatge. Bibliografia bàsica (suport paper i electrònic)

- Quatieri, T. F. 2001. Discrete-Time Speech Signal Processing: Principles and Practice. Prentice Hall.
- Rabiner, L. R. and R. W. Schafer. 2007. Introduction to Digital Speech Processing. Foundations and Trends in Signals Processing, Vol. 1, Nos. 1-2, 2007.

6.2. Fonts d'informació per a l'aprenentatge. Bibliografia complementària (suport paper i electrònic)

- Rabiner, L. R. and R. W. Schafer. 1978. Digital Signal Processing of Speech Signals. Prentice Hall.
- O'Shaughnessy, D. 1999. Speech communications: human and machine. Wiley, John & Sons.
- Rabiner, L. R. and B. H. Juang. 1993. Fundamentals of Speech Recognition. Prentice Hall.
- Park, Sung-won. Linear Predictive Speech Processing.
- Park, Sung-won. Discrete Wavelet Transform.
- Spanias, Andreas. 1994. "Speech Coding: A Tutorial Review". Proceedings of the IEEE.
- Pan, Davis. 1995. "A Tutorial on MPEG/Audio Compression". IEEE Multimedia Journal.
- Rabiner, Lawrence. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". Proceedings of the IEEE.

6.3. Recursos didàctics. Material docent de l'assignatura

- Per a cada sessió de teoria hi ha disponible uns apunts a la Web de l'assignatura.
- Per a cada sessió de seminari hi ha disponible una col·lecció d'activitats a la Web de l'assignatura.
- Per a cada sessió de pràctiques hi ha disponible el enunciat de la pràctica a la Web de l'assignatura.

6.4. Programari

- PRAAT <http://www.fon.hum.uva.nl/praat/>
- Octave <http://www.gnu.org/software/octave/>
- MATLAB

7. Metodologia

7.1. Enfocament metodològic de l'assignatura

El procés habitual d'aprenentatge per cada un dels blocs de continguts està compost per una sessió de teoria, un seminari i una part d'una sessió de laboratori. Cada bloc comença amb una sessió de teoria en la que es presenten certs fonaments teórico-pràctics. Aquesta activitat es realitza en grup gran. L'estudiant ha de complementar aquesta activitat amb una lectura detinguda dels seus propis apunts i del material addicional que el professor hagi proporcionat. Per exemple, una sessió de teoria de 2 hores, convenientment aprofitades, requerirà un treball addicional fora de l'aula d'1 hora per part de l'estudiant.

Posteriorment es realitzarà un seminari centrat en la resolució d'exercicis o problemes per posar en pràctica els conceptes i tècniques presentades a la sessió de teoria. Per als primers exercicis de la sessió es proporcionaran les solucions, però per a la resta no. L'objectiu és que l'estudiant consolidi els fonaments per tal que posteriorment pugui resoldre problemes de major complexitat. Aquesta activitat es realitza en grups petits de seminari on tots els estudiants han de participar activament en la resolució dels problemes.

El següent pas en el procés d'aprenentatge és el laboratori o sessió de pràctiques. En ell es proposen uns problemes pràctics que requereixen un disseny previ de la solució a implementar i que han d'integrar diferents conceptes i tècniques. Aquesta activitat es pot realitzar per parelles, i es pressuposa que continua fora de l'aula.

L'últim pas en el procés d'aprenentatge de cada bloc de continguts és el de la realització de l'examen final per comprovar si l'estudiant ha adquirit les competències demanades.

Set.	Blocs de contingut	Sessions presencials			Hores d'estudi
		Grup gran (2h)	Grup petit (1h)	Grup mitjà (2h)	
1	1. Introducció 2. Fonaments	T1			4
2	2. Fonaments	T2	S1. Acústica		5
3	3. Anàlisi espectral	T3	S2. Anàlisi espectral	L1. Enregistrament i Anàlisi	5
4	4. Models perceptuals	T4		L2. Model espectral	5
5	5. Models producció	T5	S3. Models	L3. Anàlisi i síntesi LPC	5
6	6. Codificació	T6	S4. Codificació		7
7	7. <i>Text-To-Speech Synthesis</i>	T7	S5. Control		6
8	8. Reconeixement	T8	S6. Reconeixement	L4. Cepstrum	6
9	8. Reconeixement	T9	S7. Transformacions	L5. Reconeixement	6
10			S8. Repàs. Arees de treball		8
	Preparació de l'examen final				7
	Total	18	8	10	64

Teoria: es corresponen amb els blocs de continguts esmentats. 18 hores (9 sessions de 2 hores).

- Teoria 1: Introducció.
- Teoria 2: Fonaments.
- Teoria 3: Anàlisi espectral.
- Teoria 4: Models perceptuals.
- Teoria 5: Models basats en la producció de veu.
- Teoria 6: Codificació.
- Teoria 7: *Text-To-Speech Synthesis*.
- Teoria 8: Reconeixement de parla: Anàlisi cepstrum, Sistemes de reconeixement automàtic.
- Teoria 9: Reconeixement de parla: Models ocults de markov.

Seminaris: 8 sessions de 1 hora.

- Seminari 1: Acústica de la veu.
- Seminari 2: Anàlisi espectral.
- Seminari 3: Models de parla.
- Seminari 4: Codificació de la parla.
- Seminari 5: Control de coneixements (T1-T6).
- Seminari 6: Reconeixement de la parla.
- Seminari 7: Transformacions de la parla.

- Seminari 8: Revisió de coneixements. Àrees de treball relacionades amb el processament de la veu.

Laboratoris: 5 sessions de 2 hores.

- Laboratori 1: Enregistrament i anàlisi espectral de veu.
- Laboratori 2: Model espectral.
- Laboratori 3: Anàlisi i síntesi de veu.
- Laboratori 4: Anàlisi Cepstrum.
- Laboratori 5: Reconeixement de la Parla.

8. Programació d'activitats

Programació de sessions presencials

Professors: Emilia Gómez, **Waldo Nogueira**, Martí Umbert

	Dimarts 14.30-16.30	Dijous 16.30-18.30	Divendres 18.30-20.30
1 9 - 13 abr	Lecture 1: Acoustics and Speech Processing	Lecture 2: Production and Classification of Speech Sounds	Seminar 1: Voice Acoustics 18:30-G3, 19:30-G4
2 16 - 20 abr	Seminar 1: Voice Acoustics 16:30-G1, 17:30-G2		
3 23 - 27 abr	Lecture 3: Short-time Analysis of Speech Signals	Seminar 2: Spectral Analysis 16:30-G1, 17:30-G2	Seminar 2: Spectral Analysis 18:30-G3, 19:30-G4
4 30 abr-4 maig	FESTIU	16:30 Lab 1: Speech Analysis-G2	18:30 Lab 1: Speech Analysis-G1
5 7 - 11 maig	Lecture 4: Hearing, auditory perception and perceptual models. Sinusoidal Models.	16:30 Lab 2: Sinusoidal Models-G1	18:30 Lab 2: Sinusoidal Models-G2
6 14 - 18 maig	Lecture 5: Source-filter models	Seminar 3: LPC - 16:30-G1, 17:30-G2 16:30 Lab 3: LPC-G2	Seminar 3: LPC - 18:30-G3, 19:30-G4 18:30 Lab 3: LPC-G1
7 21 - 25 maig	Lecture 6: Speech coding	Seminar 4: Speech Coding 16:30-G1, 17:30-G2	Seminar 4: Speech Coding 18:30-G3, 19:30-G4
8 28 maig- 1 jun	Lecture 7: Text-to-Speech Synthesis	Seminar 5: Test 16:30-G1, 17:30-G2	Seminar 5: Test 18:30-G3, 19:30-G4

9 4 - 8 jun	Lecture 8: Cepstral Analysis and Automatic Speech Recognition	Seminar 6: Speech Recognition 16:30-G1, 17:30-G2 16:30 Lab 4: Cepstrum-G2	Seminar 6: Speech Recognition 18:30-G3, 19:30-G4 18:30 Lab 4: Cepstrum-G1
10 11 -15 jun	Lecture 9: HMMs for Speech Processing	Seminar 7: Transformations 16:30-G1, 17:30-G2 16:30 Lab 5: Speech Recognition-G2	Seminar 7: Transformations 16:30-G3, 17:30-G4 18:30 Lab 5: Speech Recognition-G1
11 18 -22 jun	Seminar 8: Review, research areas 14:30-G1/G3, 15:30-G2/G4	EXÀMENS	EXÀMENS

- Llistat d'activitats (amb termini de lliurament i avaluació)

Activitat	Data enunciat	Data lliurament	Data de lliurament de resultats
Laboratori 1	30 d'Abril	11 de Maig	18 de Maig
Laboratori 2	7 de Maig	18 de Maig	25 de Maig
Laboratori 3	14 de Maig	25 de Maig	1 de Juny
Laboratori 4	4 de Juny	15 de Juny	22 de Juny
Laboratori 5	11 de Juny	19 de Juny	22 de Juny
Control	25 de Maig	25 de Maig	3 de Juny