



UNIVERSIDAD NACIONAL DEL LITORAL

**Facultad de Ingeniería
y Ciencias Hídricas**

**Instituto de Desarrollo Tecnológico
para la Industria Química**

**Carrera de Posgrado: DOCTORADO EN INGENIERÍA
Curso: Reconocimiento automático del habla – Año Académico: 2015**

1. Objetivos.

Los objetivos de este curso introductorio al reconocimiento automático del habla son que el alumno:

- Conozca los fundamentos teóricos de las etapas de extracción de características, modelado acústico y modelado estadístico del lenguaje.
- Aprenda las bases teóricas y prácticas para la construcción de sistemas de traducción del mensaje hablado a texto, para habla continua en español.
- Domine el desarrollo de sistemas basados en modelos ocultos de Markov, para su aplicación a otros problemas de reconocimiento de patrones.
- Desarrolle su habilidad para la lectura crítica de publicaciones científicas actuales sobre el tema.

2. Programa Analítico.

I. Introducción. Estructura de los sistemas de reconocimiento del habla. Principales dificultades en la tarea de reconocimiento del habla. Tipos de reconocedores. Evolución histórica. Estado actual del conocimiento. Corpus de habla y ruido. Otras aplicaciones relacionadas.

II. Producción y percepción humana del habla. El modelo de la comunicación oral humana. El aparato fonador: estructuras, procesos y modelos de generación de la voz. Principales características acústicas de la voz. Niveles de organización estructural del habla. Nociones básicas de fonética del castellano. El sistema auditivo: estructuras, procesos y modelos de percepción. Integración en escala de mel.

III. Técnicas clásicas para el análisis de señales de voz. Análisis básico de energía por bandas. Métodos de cuantización vectorial. Modelos de predicción lineal. Coeficientes cepstrales. Coeficientes en escala de mel. Coeficientes de velocidad y aceleración. Otros enfoques: RASTA PLP, características basadas en modelos auditivos. Métodos básicos para la extracción de la frecuencia fundamental y las formantes.

Universidad Nacional del Litoral
Facultad de Ingeniería y
Ciencias Hídricas

Secretaría de Posgrado

Ciudad Universitaria
C.C. 217
Ruta Nacional Nº 168 – Km. 472,4
(3000) Santa Fe
Tel: (54) (0342) 4575-234/244 – int. 103
Fax: (54) (0342) 4575 224
E-mail: posgrado@fich.unl.edu.ar



IV. Modelado acústico. Técnicas clásicas de comparación de patrones y alineamiento temporal. Autómatas finitos. Modelos ocultos de Markov discretos, continuos y semi-continuos: definiciones, algoritmo de Viterbi, reestimación de Baum-Welch, métodos de inicialización. Enlazado de parámetros. Modelos para palabras aisladas y discurso continuo. Modelos de trifonos. Introducción al entrenamiento discriminativo. Métodos basados en redes neuronales.

V. Modelado estadístico del lenguaje y modelos compuestos para el reconocimiento de habla continua. Modelos de n-gramas. Métodos de suavizado: lineal óptimo, deleted interpolation, backing-off. Redes de palabras. Alternativas al modelado estadístico.

VI. Robustez en el reconocimiento del habla. Métodos basados en realce y filtrado. Métodos de proyección en el dominio de las características. Métodos basados en la adaptación de los modelos acústicos. Métodos de adaptación al hablante.

VII. Optimización de representaciones. Wrappers para selección de características. Optimización de bancos de filtros. Codificación mediante splines. Selección de características basadas en wavelets. Adaptación dinámica de los conjuntos de datos.

VIII. Reconocimiento de emociones en señales de voz. Motivaciones del reconocimiento del estado del hablante y su aplicabilidad. Introducción al concepto de emoción, modelos y categorías. Estructuras de los sistemas de reconocimiento. Corpus disponibles y metodología de validación. Características y clasificadores más utilizados. Dificultades, motivaciones y nuevos desafíos.

3. Bibliografía.

Libros:

1. J. Deller, J. Proakis, J. Hansen, Discrete-Time Processing of Speech Signals, Macmillan Publishing, 1993.
2. C. M. Bishop, Pattern Recognition and Machine Learning. Springer: Information Science and Statistics, 2006.
3. X. D. Huang, Y. Ariki, M. A. Jack, Hidden Markov models for speech recognition. Edinburgh University Press, 1990.
4. F. Jelinek, Statistical methods for speech recognition. The MIT Press, 1999.
5. J. Junqua, J. Haton, Robustness in automatic speech recognition. Kluwer Academic Publishers, 1995.
6. J. Junqua, G. van Noord, Robustness in language and speech technology, Kluwer Academic Publishers, 2001.
7. B. D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, 1999.
8. R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification. Wiley-Interscience, 2001.

Universidad Nacional del Litoral
Facultad de Ingeniería y
Ciencias Hídricas

Secretaría de Posgrado

Ciudad Universitaria
C.C. 217
Ruta Nacional Nº 168 – Km. 472,4
(3000) Santa Fe
Tel: (54) (0342) 4575-234/244 – int. 103
Fax: (54) (0342) 4575 224
E-mail: posgrado@fich.unl.edu.ar



9. L. Rabiner, B-H. Juang, Fundamentals of Speech Recognition. Prentice Hall, 1993.
10. K. N. Stevens, Acoustic Phonetics. The MIT Press, 1999.
11. S. Furui, Digital Speech Processing, Synthesis, and Recognition, Marcel Dekker, 1989.
12. X. Huang, A. Acero, H-W. Hon, Spoken Language Processing: A Guide to Theory, Algorithm and System Development, Prentice Hall, 2001.

Artículos tutoriales básicos y de revisión:

1. J. W. Piccone, "Signal Modeling Techniques in Speech Recognition", Proceedings of the IEEE, Vol. 81, N°9, pp. 1215-1247, 1993.
2. Y. Gong, "Speech Recognition in Noisy Environments: A Survey", Speech Communication, Vol.16, N° 3, pp. 261-291, 1995.
3. M. J. F. Gales, "Predictive Model-Based Compensation Schemes for Robust Speech Recognition", Speech Communication, Vol. 25, N° 1-3, pp. 49-74, 1998.
4. R. P. Lippmann, "Speech Recognition by Machines and Humans", Speech Communication, Vol. 22, N° 1, pp. 1-15, 1997.
5. D. O'Shaughnessy, "Automatic Speech Recognition: History, Methods and Challenges", Pattern Recognition, Vol. 1, N° 10, pp. 2965-2979, 2008.

4. Docentes.

4.1 Docente responsable: Dr. Diego H. Milone

4.2 Docentes corresponsables: Dr. Hugo L. Rufiner, Dr. Leandro Di Persia, Dr. Enrique M. Albonoz, Dr. Leandro D. Vignolo

5. Conocimientos previos requeridos.

- Bases de probabilidad y estadística.
- Fundamentos inteligencia computacional.
- Análisis y procesamiento digital de señales.
- Lenguajes de programación.

6. Carga horaria (en horas de dictado efectivo).

6.1 Teoría: 30 horas

6.2 Coloquio y/o Práctica en el aula, laboratorio o campo: 30 horas

6.3 Total: 60 horas.

Universidad Nacional del Litoral
Facultad de Ingeniería y
Ciencias Hídricas

Secretaría de Posgrado

Ciudad Universitaria
C.C. 217
Ruta Nacional N° 168 – Km. 472,4
(3000) Santa Fe
Tel: (54) (0342) 4575-234/244 – int. 103
Fax: (54) (0342) 4575 224
E-mail: posgrado@fich.unl.edu.ar



7. Forma de evaluación.

7.1 Cantidad y tipo de exámenes parciales: defensa de trabajos prácticos en el cursado

7.2 Tipo y duración del examen final: oral (o escrito si hay más de 10 inscriptos), 3 horas

8. Fecha tentativa de inicio del dictado y duración del Curso (en semanas).

Inicio: 11 de septiembre

Duración: 10 semanas

9. Cupo de alumnos (cantidades mínima y máxima).

Sin cupo.

10. Lugar, día(s) y horario(s) de dictado y requerimientos.

10.1 Lugar: FICH

10.2 Días y horarios de dictado: viernes de 15 a 19 hs (a coordinar con los alumnos).

10.3 Requerimientos: pizarra y cañón, computadoras para las clases prácticas.

Santa Fe, 05/08/2015

Diego Milone

Universidad Nacional del Litoral
Facultad de Ingeniería y
Ciencias Hídricas

Secretaría de Posgrado

Ciudad Universitaria
C.C. 217
Ruta Nacional Nº 168 – Km. 472,4
(3000) Santa Fe
Tel: (54) (0342) 4575-234/244 – int. 103
Fax: (54) (0342) 4575 224
E-mail: posgrado@fich.unl.edu.ar