Ricardo de Córdoba Herralde

# Person-machine Dialogue Systems

2nd semester

Departamento de
Ingeniería
Electrónica

Universidad Politécnica de Madrid

Máster Oficial en Ingeniería
de Sistemas Electrónicos

www.die.upm.es

# Contenido



# Person-machine Dialogue Systems

*Number of credits: 3 theory + 1 practical*

*Semester: 2*

*Type: optional*

## Objectives

This course is devoted to the study of the various modules involved in an interaction system or of human-machine dialog. Starting with an overview on dialogue systems and their problems, to go on to address the key modules that make it up, describing its operation, the research alterna-tives adopted to achieve optimal system performance and the problems of each.

Each of the modules will be started from a basic level and go up to describing the most ad-vanced algorithms and techniques with which we will get the most robust and reliable systems.

The course is based on lectures to acquire the desired skills, but it also includes a set of applica-tion case studies, specially selected, to be solved in common and that allow the application skills to be acquired.

This will enhance the interaction with the students so they can apply the acquired knowledge in a final project of the subject.

## Program

The course will be cover the following topics:

1. Dialogue system architecture

2. Fundamentals of production and Speech perception

3. Synthesis and generation of response

4. Speech recognition: parameterization and quantification

5. Speech recognition: hidden Markov models

6. Continuous speech recognition

7. Adaptation

8. Language models

9. Speaker identification and language identification

10. Speech understanding and translation

11. Synthesis and recognition of emotions and multimodal interaction

12. HTS synthesis

13. Design methodologies and user modeling

14. Evaluation of dialogue systems

## Teachers

Coordinator: Ricardo de Córdoba Herralde

Teachers: Rubén San Segundo Hernández, Roberto Barra Chicote

## Teaching Methodology

The subject will be taught by:

• Lectures

• We carried out a project related to any of the modules described in the course

## Evaluation

Students complete the course with a final project of individual character to be presented publicly in English as part of activities to acquire transversal competences of documentation, communi-cation and publication.

The report must be presented in the typical format for IEEE conference papers (http://www.ieee.org/conferences_events/conferences/publishing/templates.html) with aim of encouraging the student, not only through the reading and interpretation of scientific and tech-nical documents, but also its correct wording.

The final project must be eminently practical, and in it should be applied some of the techniques described in the course, preferably, a problem that may be related to research or professional activity of the student.

The written report will be the 70% of the final grade. However, the teacher also will observe the ability of students to communicate effectively and concisely the technical information, knowledge, justifications, etc. and to answer the questions he may pose them. The oral presen-tation will be the 30% of the grade.

## Comunicación con el profesorado

- Ricardo de Córdoba Herralde, despacho B-108, cordoba@die.upm.es, ext 4209
- Rubén San Segundo Hernández, despacho B-109, lapiz@die.upm.es, ext. 4228
- Roberto Barra Chicote, despacho B-112, barra@die.upm.es, ext 4254

## Bibliografía

All material is made accessible through the Web page of the course well in advance of the delivery of the corresponding lectures. In this way, students have at all times appropriate material for easy tracking of classes.

We recommend the following general bibliography:


 Hidden Markov Models for Speech Recognition. X.D.Huang, J. Ariki, M. A. Jack. Edinburgh University Press, 1990.

 Spoken Language Processing, Huang, X., Acero, A., Hon, HW Prentice Hall, New Jersey, 2001.


For parameterization:


 Comparison of Parametric Representation for Monolyllabic contiuously Spoken Word Recognition in Sentences. S. B. Davis and P. Mermelstein. IEEE Transac-tions on Acoustics Speech and Signal Processing, Vol ASSP-28, No. 4, p. 357-366, Aug. 1980.

 Speaker-Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum. S. Furui. IEEE Transactions on Acoustics Speech and Signal Processing, Vol ASSP-34, n. 1. February 1986.

 Perceptual linear predictive (PLP) analysis of speech. Hermansky, H. 1990. JASA, p. From 1738 to 1752.

 Rasta-PLP speech analysis technique. Hermansky, H., N. Morgan, A. Bajja, P. Kohn. IEEE ICASSP 1992, pp.. 121-124.

 Towards handling the acoustic environment in spoken language processing. Hermansky, H., N. Morgan. ICSLP 1992, pp. 85-88.

RASTA Processing of Speech. Hermansky, H., N. Morgan. IEEE Trans. on Speech and Audio Processing, 1994, Vol 2, No. 4, p. 578-589.

For Vector Quantization:

"Vector Quantization". R.M.Gray. IEEE ASSP Magazine, April 1984.

An algorithm for vector quantization design. Yoseph Linde, Andres Buzo, and Robert M. Gray. IEEE Transactions on Communications, 28 (1) :84 - 95, Janu-ary 1980.

Efficient vector quantization using an N-path `Binary Tree Search Algorithm. San-Segundo, R., R. Cordoba, J. Ferreiros, A. Gallardo, J. Colas, J. Pastor, Y. Lopez. Eurospeech 1999, pp.. 93-96.

For Markov Models:

Isolated and Connected Word Recognition, Theory and selected applications. L. R. Rabiner. IEEE Trans on Communications, Com Vol 29, n,. 1981

An Introduction to Hidden Markov Models. L. R. Rabiner and B.H. Huang. IEEE ASSP Magazine, January 1986.

A tutorial on Hidden Markov Models and Selected Applications in Speech Rec-ognition. L.R. Rabiner. Proceedings of the IEEE, Vol 77, n. 2, February 1989.

Acoustic Modeling for Large Vocabulary Speech Recognition. C. H. Lee, L. R. Rabiner, R. Pieraccini and J. G. Wilpon. Computer Speech and Language (1990) 4, 127-165.

Improved acoustic modeling  With The SPHINX speech recognition system. Huang, X.D., K.F. Lee, H.W. Hon, M.Y. Hwang. IEEE ICASSP 1991, pp. 345-348.

semicontinuous Phoneme classification using HMMs. Huang, X.D. IEEE Trans. on Signal Processing, 1992, vol. 40, No. 5, pp. 1062-1067

A comparative study of discrete, semicontinuous and continuous HMMs. Huang, X.D., H.W. Hon, M.Y. Hwang, K. F. Lee. Computer Speech and Lan-guage, 1993, No. 7, pp.. 359-368.

Subphonetic Modeling with Markov States - senone. Hwang, M.Y., X.D. Huang. IEEE ICASSP 1992, pp.. 33-36.

Senones, Multi-Pass Search and Unified Stochastic Modelling in SPHINX-II. Hwang, M.Y., F. Alleva, X.D. Huang. Eurospeech 1993, vol. 3, pp.. From 2143 to 2146.

Improved acoustic modeling for speaker independent large vocabulary CSR. Lee, C.H., E. Giachin, L.R. Rabiner, R. Pieraccini, A. E. Rosenberg. IEEE ICASSP 1991, pp. 161-164.

Phonetic Context-Dependent HMMs for Speaker-Independent Continuous Speech Recognition. Lee, K.F. IEEE Trans. on ASSP 1990, Vol 38, n1 4, pp.. 599-609.

Large vocabulary CSR using HTK. Woodland, P.C., J.J. Odell, V. Valtchev, SJ Young. IEEE ICASSP 1994, pp.. II-125-128.

 The use of state tying in continuous speech recognition. Young, S.J., P.C. Woodland. Eurospeech 1993, pp. From 2203 to 2206.

 Different clustering strategies for distribution using discrete, semicontinuous and continuous HMMs in CSR. Córdoba, R., J. M. Pardo. ICSLP 1996, p. From 1101 to 1104.

 State Clustering Improvements for Continuous HMMs in a Spanish Large Vo-cabulary Recognition System. Córdoba, R., J. Macias-Guarasa, J. Ferreiros, JM Montero, J.M. Pardo. ICSLP 2002, p. 677-680.

 Different alternatives sharing parameters Nuos continuous HMM in speech recognition system isolated, Gavina Barroso, Da-vid, Thesis, 2000.

Adaptation of HMMs for:

 Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. Legetter, C. J., Woodland, P. C. Computer Speech and Language, 9, pp 171-185, 1995.

 Cluster Adaptive Training of Hidden Markov Models. Wales, MJF, IEEE Transactions on Speech and Audio Processing, Vol 8, No. 4, July 2000.

 The Generation and Use of Regression Class Trees for MLLR Adaptation. Ga-les, MJF, University of Cambridge, August 1996

 Maximum Likelihood Linear Transformations for HMM-based speech recogni-tion. Wales, MJF, Computer Speech and Language, 12, pp. 75-98, 1998

 Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observa-tions of Markov Chains. Gauvain, JL, Lee, CH, IEEE Transactions on Speech and Audio Processing, Vol 2, No. 2, April 1994

 Adaptive methods for speech and speaker recognition. Junqua, J.C., Kuhn, R. Tutorial of the International Conference on Spoken Language Processing (ICSLP), 2002.

Structural  Speaker Adaptation Using MAP Hierarchical Priors. Shinoda, K., Lee, C. H. Proc. IEEE Workshop on Automatic Speech Recognition and Understand-ing, p. 381-388, Santa Barbara, 1997

 Speaker Adaptation: Techniques and Challenges. Woodland, P. C. Proc. IEEE Workshop on Automatic Speech Recognition and Understanding, p. 85-90, 1999.

 Rapid Speaker Adaptation in Eigenvoice Space. Roland Kuhn, J.C. Junqua, P. Nguyen, N. Niedzielski. IEEE Transactions on speech and audio processing, Vol 8, No. 6, December 2000, p. 695-707.

 Eigenvoices Using Self-Adaptation for Large-Vocabulary Continuous Speech Recognition. P. Nguyen, L. Rigazio, R. Kuhn, Junqua J.-C., and C. Wellekens, in ISCA ITR Workshop on Adaptation Methods for Speech Recognition, pp. 37-40, 2001.

- Improved Recognition Using Cross-Task MMIE Training. Cordoba, R., P.C. Woodland & M.J.F. Wales. IEEE ICASSP 2002, pp. 85-88.

- Study of Speaker Adaptation Techniques in Speech Recognition Systems, Diaz, Sergio, Project Thesis, UPM, 2003.

- Cross-Task Adaptation and Speaker Adaptation in Air Traffic Control Tasks. Córdoba, R., J. Ferreiros, JM Montero, F. Fernandez, J. Macias-Guarasa, S. Diaz. Third Conference on Speech Technology, p. 93-97. November 2004.

To Identify speakers:

- Speaker Verification Using Mixture Decomposition Discrimination. R. Sukkar, M. Gandhi, and A. Setlur. IEEE Trans. SAP, Vol 8, p. 292-299, 2000.

- Speaker Verification Using Adapted Gaussian Mixture Models. D. A. Reynolds, T.F. Quatieri, and R. B. Dunn. Digital Signal Processing Review Journal, Janu-ary 2000.

- Speaker verification over the telephone. L. F. Lamel, J. L. Gauvain. Speech Communication 31 (2000) 141-154.

- Speaker-specific mapping for text-independent speaker recognition. H. Misra, S. Ikbal, B. Yegnanarayana. Speech Communication 39 (2003) p. 301-310.

- Robustness to telephone handset distortion in speaker recognition by dis-criminative feature design. Larry P. Heck, Yochai Konig, M. Kemal Sonmez, Mitch Weintraub. Speech Communication 31 (2000) 181-192.

- SMOKY: A large speech corpus in Spanish for speaker characterization and identification. J. Ortega-Garcia, J. Gonzalez-Rodriguez, V. Marrero-Aguiar. Speech Communication 31 (2000) 255-264.

- Jin, Q., Schultz, T., Waibel, A., "Phonetic Speaker Identification", ICSLP 2002, p. From 1345 to 1348.

For language recognition:

- Zissman, MA, "Comparison of four Approaches to automatic language identi-fication of telephone speech," IEEE Trans. Speech and Audio Processing, vol. 4 (1), p. 31-44, 1996.

- Torres-Carrasquillo, PA, Reynolds, DA, Deller Jr., JR, "Language identification using Gaussian mixture model tion tokenization", IEEE ICASSP 2002, pp. I-757-760.

- Wong, E., Sridharan, S., "Methods to Improve Gaussian Mixture Model Based Language Identification System", ICSLP 2002, p. 93-96.

- Navratil, J. 2001. "Spoken Language Recognition - A Step Toward Multilin-guality in Speech Processing". IEEE Transactions on Speech and Audio Proc-essing, Vol 9, No. 6, September. 2001, pp. 678-685.

Gauvain, J. L., A. Messaoudi, H. Schwenk. 2004. "Language Recognition using Phone Lattices". ICSLP, pp. I-25-28.

Ramasubramaniam, V., A.K.V. Sai Jayram, T.V. Sreenivas. 2003. "Language Identification using Parallel Phone Recognition". Workshop on Spoken Lan-guage Processing, India.

PPRLM Optimization for Language Identification in Air Traffic Control Tasks. Córdoba, R., G. Prime, J. Macias-Guarasa, JM Montero, J. Ferreiros, JM Pardo, Eurospeech 2003, pp. From 2685 to 2688.

For Speech Recognition connected:

The Application of Dynamic Programming to Connected Speech Recognition

Silverman, Harvey F. and Morgan, David P. IEEE ASSP Magazine, July 1990

Progress in Dynamic Programming Search for LVCSR. Ney, Hermann and Ort-manns, Stefan. Proceedings of the IEEE, vol. 88, No. 8, August 2000

Dynamic Programming Search for Continuous Speech Recognition. Ney, Hermann and Ortmanns, Stefan. IEEE Signal Processing Magazine, vol 16, n º 5. September 1999

The Use of a One-Stage Dynamic Programming for Connected Word Recognition Algoritm. Ney, Hermann. IEEE Transactions on Acoustics, Speech and Sig-nal Processing, Vol ASSP-32, No. 2. April 1984

An algorithm for Connected Word Recognition. Bridle, John S., Brown, Michael D. and Chamberlain, Richard M. Something IEEE. 1982

Connected Digit Recognition Using a Level-Building DTW Algorithm. Myers, Cory S. and Rabiner, Lawrence R. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol ASSP-29, No. 3. June 1981

Speaker Independent Connected Word Recognition Using a Syntax-Directed Dynamic Programming Procedure. Myers, Cory S. and Levinson, Stephen E.. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol ASSP-30, No. 4. August 1982

Dynamic Programming Parsing for Context-Free Grammars in Continuous Speech Recognition. Ney, Hermann. IEEE Transactions on Signal Processing, Vol 29, No. 2. February 1991

Two-Level DP-Matching - A Dynamic Programming-Based Pattern Matching Algorithm for Connected Word Recognition. Sakoe, Hiroaki. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol ASSP-27, No. 6. December 1979

An Investigation of the Use of Dynamic Time Warping for Word Spotting and Connected Word Recognition. Myers, C.S., Rabiner, L.R. and Rosenberg, S.A. Something IEEE. 1980

New DP Matching Algorithms for Connected Word Recognition. Watari, Ma-sao. ICASSP 96, pp. 1113-1116. Tokyo.

- Bellman, R. Dynamic Programming and Modern Control Theory. Academic Press, 1965

To Architectures for recognition:

- architectures and methods in speech recognition systems for large vo-cabulary. Javier Macias Guarasa. Doctoral Thesis. ETSIT-UPM. 2001

- Spoken Language Processing. Xuedong Huang, Alex Acero and Hsiao-Wuen Hon Prentice Hall PTR. 2001

For Models Language:

- Speech and Language Processing. D. Jurafsky and J. H. Martin. Prentice Hall, 2000

- Foundations of Statistical NLP. C. Manning and H. Schütze). MIT Press. 1999

- Natural Language Understanding. Allen, James. Benjamin / Cummings Publish-ing Co., Inc. 1995

- Statistical Language Modeling Using The CMU / Cambridge Toolkit. P. Clarkson and R. Rosenfeld. Eurospeech 1997

- Progress in Dynamic Programming Search for LVCSR. Ney, Hermann and Ort-manns, Stefan. Proceedings of the IEEE, vol. 88, No. 8, August 2000

- A Bit of Progress in Language Modeling. Extended Version. Joshua T. Goodman. Microsoft Technical Report MSR-TR-2001-72

- Estimation of Probabilities from Sparse Data for the Language Model Compo-nent of a Speech Recognizer. S. M. Katz. IEEE Transactions on Acoustics Speech and Signal Processing, 35 (3), p. 400-401. 1987

- Improved Backing off for n-gram Language Modeling. R Kneser and H Ney. ICASSP 1995

- Dynamic Programming Parsing for Context-Free Grammars in Continuous Speech Recognition. Ney, Hermann. IEEE Transactions on Signal Processing, Vol 29, No. 2. February 1991

- Speaker Independent Connected Word Recognition Using a Syntax-Directed Dynamic Programming Procedure. Myers, Cory S. and Levinson, Stephen E. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol ASSP-30, No. 4. August 1982.

- An Overview of Statistical Language Model Adaptation. J. Bellegarda, in ISCA ITR Workshop on "Adaptation Methods for Speech Recognition", p. 165-174, 2001.

- Two Decades of Statistical Language Modeling: Where Do We Go From Here? R. Rosenfeld, Proceedings of the IEEE, Vol 88, no. 8, 2000.

For dialogue management:

Lamel, L., Rosset, S., Gauvain, JL, Bennacef, S., Garnier-Rizet, H., Prouts, B., 2000. The LIMSI ARISE system. Speech Communication. Vol 31, No 4 pp 339-355, 2000.

Pellom, B., Ward, W., Sameer Pradhan, 2000. The CU Communicator: An Ar-chitecture for Dialogue Systems. Proc. ICSLP, Beijing, China. Vol II. pp723-726. 2000.

Rudnicky, A., Bennett, C., Black, AW, Chotomongcol, A., Lenzo, K., Oh, A., 2000. Task and domain specific modeling in the Carnegie Mellon System Communi-cator. Proc. ICSLP, Beijing, China, in September. Vol II pp 130-133, 2000.

R. San-Segundo, J.M. Montero, J. Macias-Guarasa, J. Ferreiros and JM Pardo. Knowledge-Combining Methodology for Dialogue Design in Spoken Language Systems "International Journal of Speech Technology". ISSN 1381-2416. Vol 8, issue 1, pp. 45-66. January 2005.

W. Ward, B. Pellom 1999. The CU Communicator System. Proc. IEEE Work-shop on Automatic Speech Recognition and Understanding (ASRU), Keystone Colorado.

Zue, V., 1997a. Conversational interfaces: advances and challenges. Proc. Eurospeech, Rhodes, Greece. kn-kn-9-18. 1997.

For evaluation of dialogue systems:

Charfuelán, A.M., 2004. Evaluation Techniques Dialogue Systems. Doctoral Thesis. Dept SSR. ETSIT-UPM. 2004.

DARPA Communicator. 2002. http://communicator.sourceforge.net/

DISC 99. Dialogue Engineering Best Practice Methodology. http://www.disc2.dk. 1999.

EAGLES 96. Expert Advisory Group on Language Engineering Standards. http://www.spectrum.uni-bielefeld/EAGLES/.

ELSE 99. Evaluation in Language and Speech Engineering. http://m17.limsi.fr/TLP/ELSE

E-MATER. E-Mail Access through the Telephone Using Speech Tecnology Re-sources: http://www.ub.es/gilcub/e-matter.

Walker, M.A., Kamm, C.A., Litman, D.J., 2000. Towards generally develop developing models of usability with PARADISE. Natural Language Engineering: Special Is-sue on Best Practice in Spoken Dialogue Systems, 2000.

Walker, MA, Rudnicky, A., Prasad, R., Aberdeen, J., Owen Bratt, E., Garo-folo, J., Hastie, H., Le, A., Pellom, B., Potamianos, A., Passonneau, R., Rou-kos, S., Sanders, G., Seneff, S., Stallard, D., 2001a. DARPA Communicator: Cross-system results for the 2001 Evaluation. ICSLP 2002. Vol.1, pp 269-272. Denver, CO USA, September. 2002.

MATERIAL RESOURCES AVAILABLE

The course itself does not currently have a dedicated laboratory equipped with work places in which to implement the techniques introduced. But it does provide trainees with suitable information on possible SW resources that may be available online (open-source software

licensed under GNU-GPL). Some examples of tools related to the tech-niques described in the subject might be:

− Praat (http://www.praat.org) tool developed by Paul Boersma and David Ween-ink of the University of Amsterdam, which allows the extraction of acoustic fea-tures.

− HTK (http://htk.eng.cam.ac.uk/ ) is a toolkit for estimating and using hidden Markov models.