Jornada de Seguimiento de Proyectos, 2004
Programa Nacional de Tecnologías Informáticas

DIHANA: Sistema de diálogo para el acceso a la información mediante habla espontánea en diferentes entornos TIC2002-04103-C03

José Miguel Benedí *
Universidad Politécnica de Valencia

Amparo Varona †
Universidad del País Vasco

Eduardo Lleida ‡
Universidad de Zaragoza

Abstract

This project has been undertaken to design and develop a dialogue system for information access using spontaneous speech in different environments. The project especially favors design options that improve the robustness of the system under real conditions. The project studies the use of spontaneous speech using different telephone channels and in non-telephone environments like automobiles. The goal is to extend the knowledge of spontaneous speech and to propose new methodologies applied to language modeling, speech understanding and dialogue management. The project will culminate with the development and evaluation of final prototype. The project includes the production of dialogue corpora in Spanish, which are needed for the development of different modules. These corpora, which will be made public to the scientific community at the end of the project, constitutes an important added value.
Keywords: dialogue, speech understanding, spontaneous speech, robust speech recognition and speech enhancement.

1 Project objectives

The main goal of this project is to study and develop a robust dialogue system for information access using spontaneous speech in different environments. The objectives are both scientific and technological, in nature:

Scientific objectives

- To propose new speech processing and acoustic modeling techniques to improve the robustness of speech recognition systems in adverse environments, particularly in car environments.

*Email: jbenedi@dsic.upv.es
†Email: amparo@we.lc.ehu.es
‡Email: lleida@unizar.es

- To investigate spontaneous speech events and their integration at the acoustic, lexical and syntactic levels of the dialogue system.
- To make contributions in the area of language modeling.
- to develop new understanding models based on machine learning techniques of stochastic finite state and neural network models.
- To propose a new stochastic framework to develop dialogue modules in a spoken language dialogue system.

Technological objectives

- To acquire and annotate large spoken dialogue corpora in Spanish.
- To develop a spoken language dialogue system that integrate both technological and scientific results y a modular architecture that uses a client/server architecture.

To achieve these objectives, the project has been developed with the joint effort of three research groups: *Universidad de Zaragoza* (UZ), *Universidad del País Vasco* (EHU) and *Universidad Politécnica de Valencia* (UPV). The UZ group comes from the signal processing and communications areas and provides fundamental knowledge to apply speech parameterization and acoustic modeling techniques. The EHU and UPV groups come from the pattern recognition area and have knowledge on machine learning techniques to study spontaneous speech events and to develop language models, and semantic and dialogue modules.

The DIHANA project started on December, 2002 and will be completed in thirteen months. The work packages has been carried out according to the timetable described in the Technical Annex. According to this timetable most of the tasks which have been developed to date are related to acquiring and annotating the corpora and to developing the initial prototype of the dialogue system. For the remaining part of the project, the following tasks are scheduled to be completed: the estimation of the different stochastic models of the dialogue system with the new corpus; the integration of the new versions of the different models in the dialogue system; and the evaluation of the final dialogue system.

2 Level of success

This section will review the main achievements of the project following the work package distribution descibed in the Technical Annex. The objectives are also related to the work package distribution.

WP0: Project coordination

The project coordination consisted principally of the following activities:

1. *Meetings*.- Six plenary meetings have been held throughout this period. Also, several other meetings have taken place for the partners to discuss and make decisions about specific issues. The minutes of every meeting have been taken and distributed.
2. WEB *site*.- All the internal reports, resources and tools produced for the project are maintained at a common web site: `http://www.dihana.upv.es`.
3. An electronic mail list was made and has been used during the project to discuss the general aspects of the project: `dihana@dsic.upv.es`.

WP1: Speech corpus acquisition

The design and acquisition of a spontaneous speech dialogue corpus in Spanish has been completed. The selected task consists of information retrieval by telephone for wide nation trains. Queries are restricted to timetables and prices for long-distance trains. A total of 900 dialogues were acquired using the *Wizard of Oz* technique (225 different users from three sites). To do this, the design and planification of both the dialogue scenes and the wizard strategy were done. Moreover, a new server module, *Wizard of Oz*, was added to the DIHANA distributed dialog architecture. The *Wizard of Oz* server module has replaced the dialog management server module. This new server module allows the *Wizard of Oz* to control the entire dialog system, to listen the user, to receive the output of the speech recognition and understanding modules, and to send speech outputs to the user.

The corpus is composed by 900 dialogues from 225 users (153 males and 72 females) with 6278 user turns and 9129 wizard turns. There is an average of 7 user turns and 10 wizard turns per dialogue. The average number of words per user turn is 7.74. A total of 10.8 hours of speech by the users was acquired. To improve the acoustic models, two additional corpora were acquired. Each speaker read 16 different sentences (8 referred to the task and 8 phonetically balanced sentences); that is, 1800 sentences per corpora.

Acquisition of a multimodal car database.- This task is focused on the acquisition of a multichannel multimodal database, named `AV@CAR`, for automatic audio-visual speech recognition in cars. The purpose of this corpus is to develop noise cancelation systems for car environments, new robust speech analysis algorithms for speech recognition, acoustic model adaptation and compensation, and new multimodal architectures for audio-visual recognition and authentication. This database is currently being recorded jointly with the Computer Vision group of the Aragon Institute of Engineering Research. The original proposal, has been extended to a multimodal database: speech and face. The database is also being recorded in two different audio-visual environments: a real car environment and a studio-like controlled environment. Twenty speakers were recorded. The database has been expanded to include three dimensional pictures of the speaker faces.

WP2: Audio Interface and Acoustic Modeling

Noise and echo cancelation, application in vehicle communication systems.- This task was been completed in the first 18 months of the project. Several algorithms for acoustic and electric echo cancelation and noise reduction have been developed and tested.

A speech reinforcement system for car cabin communication systems is being improved. The purpose of the speech reinforcement system is to increase the intelligibility of the speech signal inside the cabin of a vehicle to improve the communications between passengers. Results of this research have been applied in one international patent (jointly with Lear Co.).

Active Noise Control.- The main results that have been achieved are:

- The proposal of a new active noise-control algorithm, known as filtered-x sequential LMS (FxSLMS), which improves the convergence speed. Also, an active headrest for the driver based on a feedback control algorithm was developed using a low-cost DSP board.
- The proposal of multichannel algorithms with hybrid feedforward-feedback control strategy for laboratory environments. Algorithms such as FxLMS and FxGAL have been tested in a simulated car-like environment.

Robust speech signal analysis for noisy environments.- The research of this task is focused on the definition of new feature representations of the speech signal. Feature representations based on multi-band decomposition, and feature normalization based on minimum mean square error, histogram normalization and multi-environment models are being studied. The results with the SpeechDat Car database show an improvement of 68% in the word error rate over similar results using multi-condition training acoustic models.

Development of a speech input/output front-end with echo and noise cancelation.- Several prototypes of a speech input/output front-end with a low-cost DSP board have been built. The selected echo and noise cancelation algorithms have been used in the front-end. The system is autonomous, and could be directly installed into a vehicle. As a hands-free system, it fulfills all the echo cancelation requirements of the ITU-T G-167 recommendation for hands-free systems.

Acoustic Modeling.- Sub-phoneme models have been defined for acoustic modeling. The sub-phoneme models are composed of context-dependent models representing the left and right context of a phoneme and the context independent phoneme. A set of experiments has been performed using the BASURDE (TIC98-0423-C06) database. The minimum word error rate reached so far is 17%. The new DIHANA database is expected to reduce this error rate by retraining the acoustic models with the new data.

Acoustic model adaptation/compensation.- Classical cepstral normalization techniques in speech recognition have two important problems: the independence assumption of the feature vector components, and the mismatch error between perfect and proposed transformations. To deal with these problems, a multi-environment rotation transformation and the use of transformed space acoustic models approaches are under study. Some experiments with the SpeechDat Car database were carried out in order to study the behavior of the proposed techniques. An average improvement of 83.62% in word error rate was obtained, providing better results than the one obtained using multi-condition training acoustic models.

WP3: Spontaneous speech

Speech corpus labeling.- Spontaneous speech phenomena have been labeled from an acoustic, lexical and syntactic point of view, following a well-established annotation scheme. Labeled spontaneous speech events include non-speech acoustic events (noises, filled pauses, etc.), events that distort the lexical content (cut off and mispronounced words), events that affect the grammatical structure of the utterances (speech repairs), and also discourse markers.

Acoustic, lexical and syntactic modeling of spontaneous speech.- A set of 13 acoustic units corresponding to acoustic events occurring in spontaneous speech has been defined and evaluated. This set has been added to the set of 23 phone-like units previously defined in Spanish. The experimental evaluation of this proposal showed that an explicit modeling of acoustic disfluencies can lead to a significant improvement in system performance, when measured in terms of phone-error rates and word-error rates. Other acoustic modelizations have been tested to improve the results: explicit modeling of duration and speaker-dependent models. For the processing of lexical disfluencies, the automatic generation of lexical graphs has been proposed and is being evaluated. For the processing of the syntactic disfluencies, specific language models will be developed and integrated into the complete system.

WP4: Language modeling

Language modeling of categories integrated as a single network.- First, we have developed a classification of all words of the application into categories. There are basically two approaches to carry out: the first approach is based on the use of the knowledge applied by expert linguists, and the second is based on the use of automatic techniques using statistical methods. We have developed category language models using both approaches. Better results were obtained using the statistical methods. Another approach that takes into account the high number of spontaneous speech phenomena that have been classified in categories is currently being developed. Main categories such as cities, hours, dates, etc. have been extracted in a semiautomatic way. In this approach there is not a complete classification of words into categories, so that a combination of word language models of and a bit set of categories can be integrated into the single network. The theoretical details of this integration are being developed in parallel with the implementation.
Generation and evaluation of language models of the task.- All the user turns of the 900 dialogues are used to obtain the language models. However, the recorded material is too small. During the dialogue acquisition process many of them were rejected for different reasons. However, we have decided to take advantage of all this material and it is being currently processed and labeled. We have also developed software to automatically generate new task sentences according to 33 categories.

Within the framework of this task, we are also using new hybrid language models that allow us to improve our classical language models. In this work, the hybrid language model is defined as a combination of a word-based n-gram, which is used to capture the local relations between words, and a category-based Stochastic Context-Free Grammar with a word distribution in categories, which is defined to represent the long-term relations between these categories.

WP5: Understanding Module

The main scientific objectives achieved in this work package are:

- The development of a smoothing technique for language modeling. This technique has been applied to the different stochastic models used in the understanding module.
- The development of understanding modules are based on different automatic learning techniques: two-level stochastic models that use N-grams and grammatical inference techniques, and stochastic N-multigram models.
- The incorporation of knowledge from the dialog manager, that is, the use of specific understanding modules depending on the dialog act.
- The incorporation of the confidence measures provided by the recognition module to the understanding module.
- The definition of two more confidence measures provided by the understanding module to the dialog manager in order to improve the detection and management of errors in the dialog system.

The main technological objectives achieved are:

- The integration of the understanding module with the recognition module in a prototype-0 in order to facilitate the acquisition task.
- The integration of the understanding module with the recognition module and the dialog manager in a prototype-1.

WP6: Dialogue Module

Semi-automatic corpus labeling.- In this task, the following works have been carried out: the definition of the rules for tagging the dialogues with dialogue acts; the design and implementation of an assisted tool for manual tagging; the design and implementation of tools for automatic tagging; and the definition and distribution of the corpus to be manually tagged.
Dialogue manager.- First, an approach to the development of the dialogue manager based on stochastic models has been explored. An important aspect of this approach is the combination of the stochastic model, which represents the probabilities of sequences of dialogue acts, with the information supplied by the user during the dialogue. We have also studied how the dialogue manager can use confidence measures that are supplied by the understanding modul.

The design of another dialogue manager based on rules has been explored. The design of this dialogue manager is based on the *Wizard of Oz* strategy used in the acquisition process, and it is directed by the attributes to be filled in order to generate the database query. Confidence measures given by the recognition and understanding process have also taken into account.

Second, a stochastic dialogue model has been defined. Both the estimation of the parameters of the model and the decoding algorithm have been formally stated. This stochastic dialogue model has been implemented and tested on the BASURDE corpus. The upcoming task is to train and test the model with the corpus acquired in this project.
Work on VoiceXML in DIHANA.- VoiceXML technology to develop a dialogue system has been studied for the DIHANA project. This technology has been compared with traditional methods which are based on general purpose high-level languages. The work on VoiceXML has focused on the following tasks:

- *Automatic speech recognition and natural language processing.* Several grammar formalisms have been studied in order to define the set of user sentences to be understood. Java Speech Grammar Format was chosen.
- *Dialogue management.* The interaction and confirmation strategies for the system have been studied. A mixed-initiative strategy system-user was chosen.
- *Response generation and speech synthesis.* Different types of sentences that the system needs to interact with the users was studied and the appropriate patterns were created.

WP7: System Architecture and Integration

The main objectives of this work package are: the definition of the communication model; the definition and development of programming language; and the integration and evaluation of the distributed dialog system.

The DIHANA architecture is based on the BASURDE (TIC98-0423-C06) architecture. The DIHANA architecture is composed of 7 conceptual components: an audio server (AUDIO), an automatic speech recognition server(RAH), a speech understanding server(CH), a dialogue management server (GD), an oral answer generation server (GRO), a text-to-speech conversion server (CTV), and finally a communications management client (GC). Figure 1 shows the conceptual diagram. The communication between modules is performed by means of packets that are sent from one module to another using a TCP/IP protocol. All the packages of information, with the exception of the audio package, are sent from one module to another through the communications manager. The manager directs the packages towards the server destiny. The communications manager is controlled through an applet that is executed in a web browser.

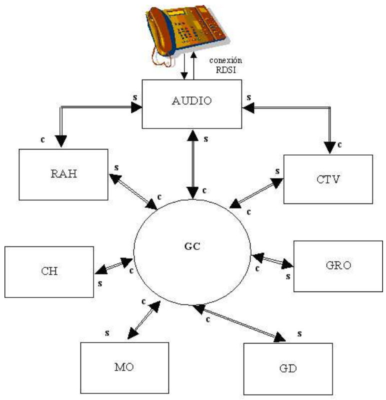

Figure 1: DIHANA architecture (c:client, s:server).

DIHANA architecture includes a *Wizard of Oz* module for human supervision of the dialog system, and the communication model has been redefined to use XML documents. The use of a XML-based communication model allows us to unify the communication protocol with the DIHANA programming language. A complete definition of the XML schema for the communication model has been released. Two libraries have been developed: a basic communication library that deals with the low-level communication between client and server, and a XML communication library that to deal with XML documents. The DIHANA programming language is currently being defined. The purpose of this programming language is to provided an easy configuration of the distributed dialog system to new applications. The programming language is a XML language that can be used to configure any of the modules of the DIHANA architecture.

3 Results

3. 1 PhD students

Universidad del País Vasco.- One PhD student with a FPI grant associated to the project is working on her PhD thesis on the task of language modeling of categories integrated as a single network. Another person was hired in September and has been included in the project budget.
Universidad de Zaragoza.- Two PhD students have worked on the project from the beginning. One of them is funded by a FPI grant associated to the project, and the other was hired the first year under contract (personal budget) and has been working under FPU grant since February. A third PhD student started to work on the project in February in the area of robust automatic speech recognition.
Universidad Politécnica de Valencia.- Two PhD students with FPI grants associated to the project are working and doing their PhD theses in the areas of understanding and dialogue. Two new PhD students have been hired on the project budget, and are working in the project and starting their PhD degree.

3. 2 Internal reports

Several reports have been produced during the development of the project and are accessible to the partners at the *web* site. These reports compile research work, conventions (in language, labeling, protocols, etc.), and resource and tool documentations. A list of these reports can be found in the Annex.

3. 3 Publications

Much of the research work of the project has been published in the main journals and conferences of the field. More than 40 publications have been compiled in the Annex. These publications are well distributed among the three research groups and represent a good balance among the different lines of research.

3. 4 Patents

The UZ research group has obtain an international patent for the speech reinforcement system built for LEAR Corporation.

3. 5 Transfer of Technology

The EHU research group is a partner of the project *GENIO, Gestor Embebido Natural de Interfaz Oral* led by *Fagor Electrodomesticos.* The main goal of this project is the development of a reduced but complete dialogue system for two different applications: the communication with household appliances and the request for subway tickets.

The UZ research group has developed a hands-free system for *Telefónica I+D* within the PROFIT project *Incorporación de las Tecnologías del Habla en el Interior de Vehículos.*

The UZ research group has developed a hands-free system for medical ambulance environment which includes a speech recognition system to control the telemonitorization system inside the ambulance. This work *Implementación y evaluación de un sistema de telemonitorización en vehículos de emergencias médicas sobre una red UMTS* was funded by *Telefónica Móviles.*

The UZ research group is also participating to improve the robustness of the speech recognition system in the project *GENIO* led by *Fagor Electrodomesticos.*

3. 6 Participation in National or International projects

The partners of this project have also participated in other related projects. A list with the most relevant ones is presented in the Annex.

3. 7 Collaboration with other groups

The UZ research group is collaborating with the Speech Technologies group of *Telefónica I+D* in the area of automotive applications. Within the Aragon Institute of Engineering Research they are also collaborating with the Computer Vision Lab in the area of biometric authentication, with the Computer Graphics Group in the area of intelligent interfaces, and with the robotics group introducing speech technologies in an autonomous wheelchair.

The UPV research group is collaborating with other national Pattern Recognition research groups, such as DIATA (Univ.Granada) coordinated by N.Perez; CV (Unv.Jaume I) coordinated by F.Pla and RFIA (Univ.Alicante) coordinated by R.Carrasco. Furthermore, the UPV is also collaborating with other international groups, such as WRTH (Unv.Aachen) group directed by H.Ney and EURISE (Unv.Saint Etienne) group directed by C.de la Higuera.

Annex

Internal Reports

- *Estrategia del Mago.* Informe técnico DIHANA TIC2002-04103. Valencia, julio 2004.
- *Adquisición del corpus de* DIHANA. Informe Técnico Proyecto DIHANA TIC2002-04103. Valencia, junio 2004.
- *Librería de comunicaciones basada en paquetes XML: Dihana v2.0.* Informe Técnico Proyecto DIHANA TIC2002-04103. Zaragoza, 2004.
- *Normas de etiquetado de diálogo para el proyecto Dihana.* Informe Técnico Proyecto DIHANA TIC2002-04103. Valencia, junio 2004.

Publications (Journals, Conference papers and book Chapters)

- L.J.Rodríguez, I.Torres, A.Varona: *Anotación de disfluencias en un corpus de habla espontánea no específico.* Actas de las II jornadas en Tecnologías del Habla. pp-17. Granada, Diciembre de 2002.
- M.J.Castro, E.Sanchis: *A Simple Connectionist Approach to Language Understanding in a Dialogue System.* Advances in Artificial Intelligence (IBERAMIA 2002), F.J.Garijo, J.C.Riquelme, M.Toro (Eds.). Springer-Verlag, LNAI Vol.2527, pp.664–673, 2002.
- J.A.Gómez, M.J.Castro, E.Sanchís: *Construcción de grafos de fonemas para un sistema de RAH desacoplado.* Actas de las II Jornadas en Tecnologías del Habla. Granada, diciembre 2002.
- E.Sanchis, M.J.Castro: *Dialogue Act Connectionist Detection in a Spoken Dialogue System.* En "Soft Computing Systems. Design, Management and Applications", A.Abraham, J.Ruiz, M.Köppen (eds.). IOS Press, Vol.87, pp.644–651. II International Conference on Hybrid Intelligent System, Santiago de Chile, diciembre 2002.
- D.Vilar, M.J.Castro, E.Sanchis: *Comparación de métodos de detección de actos de diálogo.* Actas de las II Jornadas en Tecnologías del Habla. Granada, diciembre 2002.
- P.Torres, E.Sanchis, E.Segarra: *El uso de medidas de confianza en un gestor de diálogo dirigido por la semántica.* Actas de las II Jornadas en Tecnologías del Habla. Granada, diciembre 2002.
- A. Casillas, A.Varona, I.Torres: *Experiments with Linguistic Categories for Language Model Optimization.* Procc of CICLing2003. Lecture Notes in Computer Science. Pp.512-515. Febrero 2003.
- L.J.Rodríguez, I.Torres: *Annotation and analysis of acoustic and lexical events in a generic corpus of spontaneous speech Spanish.* Procc of ISCA and IEEE workshop on Spontaneous Speech Processing and Recognition. pp 187-190. Tokio, Abril 2003.
- L.J.Rodríguez, I.Torres: *Comparative study of the Baum-Welch and Viterbi training Algorithms applied to read and spontaneous speech recognition.* Procc of the First Iberian conference, ibPRIA. pp 847-857. Mallorca. Mayo 2003.
- A.Varona, I.Torres: *Integrating High an Low Smoothed LMs in a CSR system.* Procc of CIARP2003. Lecture Notes in Computer Science. Pp 236-243. Noviembre 2003.
- A.Ortega, E.Lleida, E.Masgrau: *Residual Echo Power Estimation for Speech Reinforcement Systems in Vehicles.* EUROSPEECH-03, Suiza. Septiembre 2003.
- P.Ramos, A.López, A.Salinas, E.Masgrau: *Performance comparison between the FxLMS and the FxHLMS algorithms in an Active Noise Control system.* Active Noise and Vibration Control Methods, Polonia. Marzo 2003.
- L.Vicente, E.Masgrau, J.M.Sebastián: *Active Noise Control Experimental Results with FxGAL Algorithm.* InterNoise-2003, Corea. Agosto 2003.
- A.Miguel, M.I.Galiano, R.Granell, LL.F.Hurtado, J.A.Sánchez, E.Sanchis: *La plataforma de adquisición de diálogos en el proyecto DIHANA.* En Proceedings of the SEPLN: XIX Congreso de

la Sociedad Española para el Procesamiento del Lenguaje Natural, páginas 343-344. Septiembre 2003.

- DIHANA*: sistema de diálogo para el acceso a la información mediante habla espontánea en diferentes entornos*. En el marco del "Curso de Industrias de la Lengua. Conversar con el ordenador: el procesamiento del lenguaje y del habla en los sistemas de diálogo", Fundación Duques de Soria. Soria, julio 2003.
- J.García, J.A.Sánchez, J.M.Benedí: *Performance and Improvements of a Language Model based on Stochastic Context-Free Grammar*. F.Perales, A.Campilho, N.Pérez (Eds.), Springer Verlag, LNCS-2652, pp.271–278. Iberian Conference on Pattern Recognition and Image Analysis, Mallorca, junio 2003.
- D.Linares, J.M.Benedí, J.A.Sánchez: *Learning of Stochastic Context-Free Grammars by means of Estimation Algorithms and Initial Treebank Grammars*. F.Perales, A.Campilho, N.Pérez (Eds.), Springer Verlag, LNCS-2652, pp.403–410. Iberian Conference on Pattern Recognition and Image Analysis, Mallorca, junio 2003.
- D.Linares, J.M.Benedí, and J.A.Sánchez: *A hybrid language model based on stochastic context-free grammars*. In C. de la Higuera, P. Adriaans, M. van Zaanen, and J. Oncina, editors, ECML/PKDD 2003 Workshop on Learning Context-Free Grammars, pages 41-52, Septembre 2003.
- D.Linares, J.M.Benedí, and J.A.Sánchez: *Earley-based stochastic context-free grammar estimation from bracketed corpora and its use in a hybrid language model*. Procesamiento del Lenguaje Natural. ISSN: 1135-5948, N 31 , pages 183-190, Septiembre 2003.
- M.J.Castro, F.Prat: *New Directions in Connectionist Language Modeling*. J.Mira, J.Álvarez (Eds.) IWANN 2003. Springer Verlag, LNCS-2686, pp.598–605. Menorca, junio 2003.
- E.Segarra, E.Sanchis, F.García, Ll.Hurtado, I.Galiano: *Achieving full coverage of automatically learnt finite-state language models*. Workshop on Finite-State Methods in Natural Language Processing. 10th Conference of the European Chapter of the Association for Computational Linguistics. ISBN 1-932432-02-7, pp. 135-142, Budapest (Hungría), 2003.
- F.García, E.Sanchis, Ll.Hurtado, E.Segarra: *Modelos específicos de comprensión en un sistema de diálogo*. Procesamiento del Lenguaje Natural. ISSN: 1135-5948, N.31, pp.99-106, 2003.
- F.García, Ll.Hurtado, E.Sanchis, E.Segarra: *The incorporation of Confidence Measures to Language Understanding*. Lecture Notes in Aritifial Intelligence series 2807 (Springer Verlag). ISBN: 3-540-20024-X. International Conference on Text Speech and Dialogue (TSD 2003). Vol LNAI 2807, pp. 165-172, 2003.
- E.Sanchis, M.J.Castro, D.Vilar: *Stochastic Understanding Models Guided by Connectionist Dialogue Acts Detection*. En "Proceedings of 2003 IEEE Workshop on Automatic Speech Recognition and Understanding Workshop", pp 501–506, St. Thomas, U.S. Virgin Islands, diciembre 2003. IEEE.
- D.Vilar, Mª J.Castro, E.Sanchis: *Connectionist classification and specific stochastic models in the understanding process of a dialogue system*. En "Proceedings of the Eurospeech'03", volumen 1, pp 645-648, Ginebra (Suiza), septiembre 2003.
- F.Torres, E.Sanchis, E.Segarra: *Development of a stochastic dialog manager driven by semantics*. European Conference on Speech Communication and Technology. (EUROSPEECH'03). ISSN 1018-4074. pp. 605-608. (Suiza) 2003.
- M.J.Castro, D.Vilar, P.Aibar, E.Sanchis: *Clasificación de Actos de Diálogo en un Sistema de Diálogo Hablado*. En "Actas de la X Conferencia de la Asociación Española para la Inteligencia Artificial", volumen II, páginas 35-44, San Sebastián (España), noviembre 2003.
- M.J.Castro, D.Vilar, E.Sanchis, P.Aibar: *Uniclass and Multiclass Connectionist Classification of Dialogue Acts*. En "Progress in Pattern Recognition, Speech and Image Analysis", volumen 2905 de Lecture Notes in Computer Science, páginas 266–273. Springer, 2003. 8th Iberoamerican Congress on Pattern Recognition (CIARP 2003). Havana, Cuba, November 2003.

- M.J.Castro, F.Prat: *Computational Methods in Neural Modeling.* In New Directions in Connectionist Language Modeling. J.Mira and J.R.Alvarez (Ed.), Springer-Verlag. LNCS, Vol.2686, pp.598–605, 2003.
- V.Guijarrubia, I.Torres, L.J. Rodríguez: *Evaluation of a Spoken Phonetic Database in Basque Language.* Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004) Vol. 6, pp.2127-2130. Lisboa, Portugal, 2004.
- L.J.Rodríguez, I. Torres: *A speaker clustering algorithm for fast speaker adaptation in continuos speech recognition.* 7th International Conference on TEXT, SPEECH and DIALOGUE. Lecture Notes in Artificial Intelligence. Brno (Czech Republic). Setiembre 2004.
- A.Casillas, V.Fresno, M.T.González de Lena, R.Martínez: *Evaluation of Web Page Representations by Content through Clustering.* Procc of SPIRE'2004. Lecture Notes in Computer Science. pp.- 129-130. Padoba. Italia. Octubre de 2004.
- A.Varona, I.Torres: *Scaling acoustic and language model probabilities in a CSR system.* Procc of CIARP2004. Lecture Notes in Computer Science. Aceptado y pendiente de publicación. Octubre de 2004.
- A.Ortega, F.Sukno, E.Lleida, A.Frangi, A.Miguel, L.Buera: *AV@CAR: A Spanish Multichannel Multimodal Corpus for In-Vehicle Automatic Audio-Visual Speech Recognition.* Proceedings International Conference on Linguistic Resources and Evaluation, LREC-2004, Lisboa, (Portugal), Mayo 2004
- P.Ramos, R.Torrubia, A.López, A.Salinas, E.Masgrau: *Computationally Efficient Implementation of an Active Noise Control System Based on Partial Updates.* Active-2004, Williamsburg, Virginia (USA), September 2004.
- L.Buera, E.Lleida, A.Miguel, A.Ortega: *Multi-Environment models based linear normalization for speech recognition in car conditions.* in Proc. ICASSP, mayo 2004.
- L.Buera, E.Lleida, A.Miguel, A.Ortega: *Multi-Environment models based linear normalization for robust speech recognition*, in Proc. SPECOM, septiembre 2004.
- N.Alcácer, D.Griol, R.Granell, S.Grau: *Resultados del proceso de adquisición de un gran corpus de diálogo hablado en el proyecto DIANA.* Curso de Tecnologías Lingüísticas. Fundación Duques de Soria, julio 2004.
- R.López-Cózar, E.Sanchís, E.Segarra, J.M.Benedí: *Current Activities in the* DIHANA *Project.* 7th Committee Management Meeting of Spoken Language Interaction in Telecommunication (COST-278). Belgium, January 2004.
- M.J.Castro, D.Vilar, P.Aibar, E.Sanchis: *Dialogue Act Classification in a Spoken Dialogue System.* En "Current Topics in Artificial Intelligence", volumen 3040 de Lecture Notes in Artificial Intelligence, páginas 260-270. Springer-Verlag, 2004.
- R.López-Cózar, E.Sanchís, E.Segarra, J.M.Benedí: *Incorporación de VoiceXML en el Sistema de Diálogo* DIHANA. V Congreso Interacción Persona- Ordenador, Lérida, pág. 253-256. 2004.
- R.López-Cózar, R.Granell: *Sistema de Diálogo Basado en VoiceXML para Proporcionar Información de Viajes en Tren.* Procesamiento del Lenguaje Natural, vol. 33, pág. 171-178. 2004.
- D.Linares , J.M.Benedí, J.A.Sánchez: *A Hybrid Language Model based on a Combination of N-grams and Stochastic Context-Free Grammars.* ACM Trans. on Asian Language Information Processing, 3(2), pp. 113-127, 2004.
- J.M.Benedí, J.A.Sánchez: *Estimation of stochastic context-free grammars and their use as language models*, Computer Speech and Language. Accepted to be published (2005).
- A.Ortega, E.Lleida, E.Masgrau: *Speech Reinforcement System for Car Cabin Communications.* IEEE Trans. On Speech and Audio Processing, Accepted to be published (2005).

PhD Thesis

- Diego Luis Linares: *Estimación de gramáticas incontextuales estocásticas y su aplicación en un modelo de lenguaje híbrido.* PhD Thesis, Universidad Politécnica de Valencia, julio 2003. Supervisors: J.M.Benedí and J.A.Sánchez.
- Fernando García Granada: *Una aproximación estocástica para la comprensión del lenguaje.* PhD Thesis, Universidad Politécnica de Valencia, 2003. Supervisors: E.Sanchis y E.Segarra.
- Luis Javier Rodriguez: *Estudio y modelización del habla espontánea en diálogos hombre-máquina y entre personas.* PhD Thesis, Universidad del País Vasco, Julio 2004. Supervisor: M.Inés Torres.
- Lluís F.Hurtado i Oliver: *Modelos de lenguaje basados en palabras y segmentos y su aplicación a la comprensión del habla.* PhD Thesis, Universidad Politécnica de Valencia, 2004. Supervisors: E.Segarra y E.Sanchis.
- Jon Ander Gómez Adrián: *Nuevas aproximaciones al Reconocimiento Automático del Habla mediante Grafos de Palabras y técnicas de aprendizaje no supervisado.* PhD Thesis, Universidad Politécnica de Valencia, 2004. Supervisors: E.Sanchis y M^{a}J. Castro.

Links to other projects

Universidad del País Vasco

- *Subvención general a grupos de investigación.* Gobierno Vasco - UPV/EHU (9 /UPV 00224.310-13566/2001). 2002-2004.
- *Ayuda a la traducción basada en memorias de traducción.* Gobierno Vasco - INTEK . TEC-154-02. 2002-2004.
- *Desarrollo de herramientas para la creación automática y manipulaciín de memorias de traducción. Reconocimiento de nuevas unidades de traducción. Acceso y uso a través de la web.* UPV/EHU, CODE AND SYNTAX S.L. ELHUYAR KULTUR ELKARTEA (UE022/B0-2). 2002-2004.
- *GENIO Gestor Embebido Natural de Interfaz Oral.* Gobierno Vasco - INTEK . 2003-2005.
- *Diseño, adquisición y anotación de una base de datos de voz en Euskera par a el AS3200 de Scansoft.* Gobierno Vasco - 2004.

Universidad de Zaragoza

- The research group of the University of Zaragoza is a member of the BioSecure Network of Excellence (www.biosecure.info) "Biometrics for Secure Authentication". The group is contributing into the jointly executed research activities related to speech modality.

Universidad Politécnica de Valencia

- TransType2*: Computer-assisted translation.* Unión Europea IST-2001-32091. 2002-2005.
- *Sistemas híbridos para la traducción valenciano-castellano a partir de voz y texto.* CICYT (TIC2000-1599-C02-01). 2001-2003.
- *Diseño de sistemas de reconocimiento de escritura continua manuscrita y de clasificación de documentos mediante técnicas de inferencia gramatical y de reconocimiento sintáctico.* CICYT (TIC2000-1153). 2000-2003.
- *Recuperación de respuestas en Documentos Digitalizados (R2D2).* Organismo Financiador: CICYT TIC2003-07158-C04-03. (2004-2006).
- *TEFATE: Tecnologías de Estados Finitos Aplicadas a la Traducción Especializada.* CICYT TIC2003-08681-C02-02. (2004-2006).