GEOVAQA: A VOICE ACTIVATED GEOGRAPHICAL QUESTION ANSWERING SYSTEM

Jordi Luque, Daniel Ferrés, Javier Hernando, José B. Mariño and Horacio Rodríguez

Technical University of Catalonia (UPC)
TALP Research Center
Jordi Girona, 1-3, 08034 Barcelona, Spain
{aluque, javier, canton}@gps.tsc.upc.edu , {dferres,horacio}@lsi.upc.edu

ABSTRACT

In this paper we present GeoVAQA, a Restricted Domain Spoken Question Answering system in the scope of the Spanish geography. The system consists of a web-based application that allows speech input questions about Spanish geography and sends back a concise textual answer. In our system, spoken questions are recognised by an automatic speech recognition (ASR) system. We have used RAMSES, a Spanish recogniser based on semi-continuous HMMs, to perform this task. The transcribed question is the input of GeoTALP-QA, a multilingual Geographical Domain QA system. The retrieval mechanism of the QA system uses Natural Language Processing tools, a Geographical Knowledge Base and the search engine Google to get snippets. A speech question is recorded from web client and transmitted through the Internet to the demo server which transcribes the speech and retrieves a concise answer and a set of relevant snippets.

1. INTRODUCTION

The amount of information on the Internet is growing exponentially and standard word statistic-based search engines, which return a large number of irrelevant matches, are rapidly becoming obsolete. In the same way, the increasing popularity of web-based portable computing devices with limited display capabilities has created a need for advanced information access technologies that interact with the Internet using voice and other modalities (pen, gestures, haptic, and others).

Speech Recognition (SR) is the ability by which a machine identifies spoken words. Being speech the primary means of communication between humans, Automatic Speech Recognition (ASR) systems have becoming a serious alternative to classical textual human-machine interfaces through keyboard.

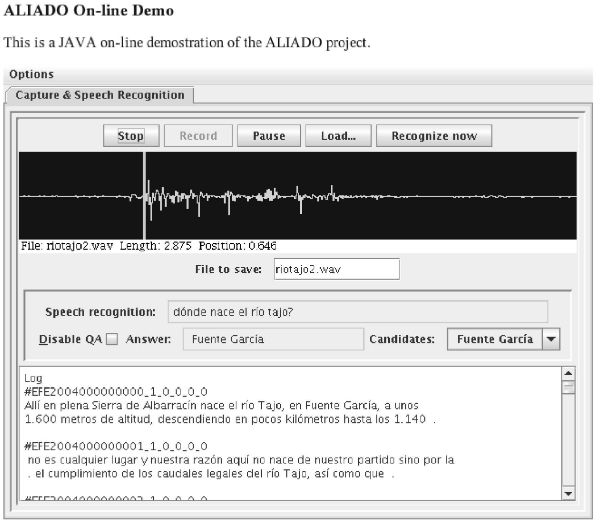
Question Answering (QA) is the task of, given a query expressed in Natural Language (NL), retrieving its correct answer. QA has become a popular task in the NL Processing (NLP) research community in the framework of different Open Domain QA evaluation contests such as: Text Retrieval Conference (TREC) and Cross-Lingual Evaluation Forum (CLEF). Despite the growing interest in Open Domain QA (ODQA), there are few systems that integrate QA and ASR with a speech input. Speech interfaces to QA systems offer significant potential for finding information with phones and mobile networked devices and represent the next generation capability for universal access by integrating state-of-the-art of QA and ASR systems. Some of them are ODQA systems. [1] presented a system with iterative refinement of QA and ASR for ODQA. They used LVCSR, a publicly available ASR from the Institute of Signal and Information Processing (ISSP)[1], trained on a database of broadcast news. [2] described a speech interface for the AnswerBus[2] QA system using a commercial dictation engine (Dragon Naturally Speaking 6.1). [3] presented QASR, a system using semantic role labelling for QA and the AT&T Watson speech recogniser. [4] presented SPIQA, a spoken interactive ODQA system for Japanese that uses a Weighted Finite-State Transducers (WFST) approach for ASR.

In this paper we describe our experiments in the adaptation and evaluation of an ASR Voice Interface for GeoTALP-QA, a multilingual Geographical Domain Question Answering (GDQA) system [5]. The basic approach of the QA module consists of applying language-dependent processes on both question and passages for getting a language independent semantic representation, and then extracting a set of Semantic Constraints (SC) for each question. Two algorithms, a Semantic Constraint Relaxing and a Frequency-based one, have been applied for answer extraction.

We outline below the organisation of the paper. In section 2, we describe the framework of the demonstration, focusing on the web strategy implementation and the user interface. In section 3, we present the overall architecture of GeoVAQA and describe briefly its main components for ASR and QA. Then, the experiments are presented in section 4. Finally, the last section provides the conclusions.

[1] **ISSP**. `http://www.isip.msstate.edu`

[2] **AnswerBus**. `http://www.answerbus.com`



ALIADO On-line Demo

This is a JAVA on-line demostration of the ALIADO project.

File: riotajo2.wav Length: 2.875 Position: 0.646

File to save: riotajo2.wav

Speech recognition: dónde nace el río tajo?

Disable QA □ Answer: Fuente García Candidates: Fuente García

```
Log
#EFE2004000000000_1_0_0_0_0
Allí en plena Sierra de Albarracín nace el río Tajo, en Fuente García, a unos
1.600 metros de altitud, descendiendo en pocos kilómetros hasta los 1.140  .

#EFE2004000000001_1_0_0_0_0
 no es cualquier lugar y nuestra razón aquí no nace de nuestro partido sino por la
 . el cumplimiento de los caudales legales del río Tajo, así como que  .
```

Fig. 1. *Applet user interface*
Demo. *http://gps-tsc.upc.edu/veu/aliado/demos.html*

2. APPLICATION DESCRIPTION

The application consists of a Geographic Restricted Domain QA system that processes oral questions in Spanish in the domain of the Spanish geography, and returns a concise answer and a set of relevant snippets. The user interface consists in a web-based applet allowing speech recording, audio transport through internet access and information about the connection status and the final answer. The applet interface was developed using Java^{TM} 2 Platform Standard Edition Development Kit 5.0 [3].

The speech signal is recorded in the user side at 16 kHz sampling frequency, quantized using 16-bit linear coding and transmitted to the server. In the server side, the ASR system performs speech recognition and returns the speech transcription to the client. At the same time, the ASR output enables question processing and the GeoTALP-QA searches the answer and the candidates. A few seconds and the QA system sends back the answer. The applet shows a concise answer, the candidate with the highest score, and the top-scored candidates. In addition, is possible to explore a brief log of snippets from Google. A screen-shot of the applet is shown in Figure 1.

The implementation strategy allows a low cost computation in the user's terminal, which is only used to record, transmit and display information. The speech recognition and the answer processing occurs in the server side. Both tasks are processed in a Intel Pentium IV 3.0 GHz getting real-time processing in the ASR task and a good response time for the QA system.

3. ARCHITECTURE SYSTEM DESCRIPTION

In Figure 2 we depict the main components and the information flow in the GeoVAQA system. Note that the question transcribed by the ASR system sometimes incorporates not only redundant information caused by the spontaneity of human speech but also irrelevant information due to recognition errors. Therefore, with this simple integration, recognition errors degrade the performance of the QA system. Some indispensable information to extract answers is deleted or substituted by other words.

3.1. RAMSES-ASR system

We have used the recognition software RAMSES [6] to perform the speech recognition. In order to obtain a good representation of the acoustic model, a multi-condition training with a large amount of data is used to represent a variety of speakers and conditions. For this purpose we employ a database recorded in real world conditions, called SPEECON [7]. We also trained the language models by presenting the application corpus of ALBAYZIN [8], a speech geographic database. The geographical vocabulary, roughly $1,100$ words, is also obtained from the application corpus of ALBAYZIN. Recognition system front-end is based on frequency filtering (FF) features [9] and includes well known robust techniques like spectral subtraction. [10] shows that FF front-end obtain the best word error rates scores in all test using SPEECON database. The back-end part of the system is based on demiphone models [11], a context dependent sub-word unit that models independently the left and the right parts of a phoneme. The demiphone simplifies the recognition system and yields a better performance than the triphone [12]. The Databases and the ASR will be described in more detail below.

3.1.1. SPEECON and ALBAYZIN databases

SPEECON database contains 600 speakers, 550 adults and 50 children. It has been collected in 5 different scenarios: office environments, public places, entertainment, car environments and children area. A detailed description of each scenario can be found in [7]. The database was recorded at 16 kHz sampling frequency and quantizes using 16-bit linear coding. Each session was recorded simultaneously by four microphones: a head-mounted close-talk microphone, a lavalier (a microphone placed just below the chin of the speaker) and two far-field conditions microphones situated at 1 and 2-3 meters from the speaker respectively. In order to obtain the speaker independent demiphone models we used a total of $20,496$ utterances from close-talk and lavalier microphone.

ALBAYZIN [8] is a Spanish spoken database designed for speech recognition purposes. We have used the application database of ALBAYZIN to obtain a

[3] http://java.sun.com/j2se/1.5.0/docs/

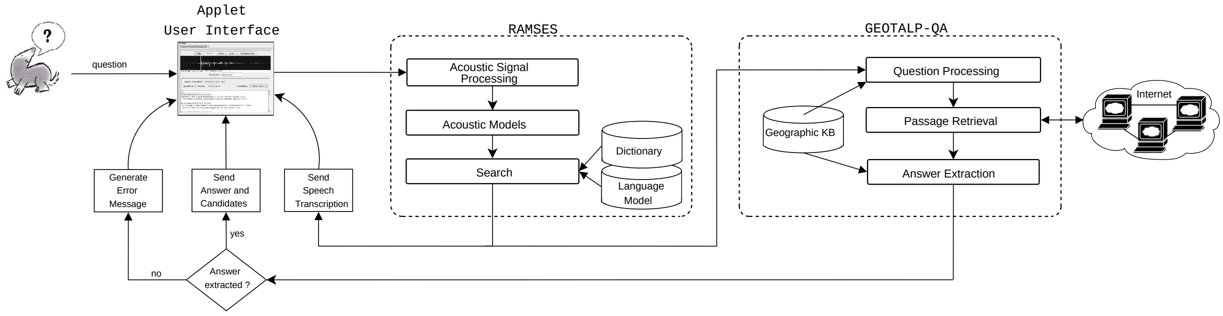


Fig. 2. *GeoVAQA System Architecture*

language model. The corpus is divided in a training set composed by $2,700$ sentences and a testing set composed by a $1,200$ sentences. This subset of the application corpus was used in several experiments in order to choice the best combination of microphones from SPEECON data. Finally, the diphonemes were trained using close-talk and lavalier subsets data. The ALBAYZIN corpus has also been divided in 78 subsets of 50 sentences each one and is uttered by 136 speakers.

3.1.2. Front-end

The front-end used obtains a signal representation by frequency filter (FF) parameters. The signal is pre-emphasized with a zero at $z = 0.97$ and Hamming windowed frames of 30 ms. were taken every 10 ms. Then, Q mel-scale filter-bank energies were computed for each frame and logarithmically compressed.

FF technique consists of a linear transformation of the log Mel-spaced filter-bank energies. That transformation essentially consists of convolving the frequency sequence of those energies with the impulse response of the FIR filter. We employ the impulse response $\{1, 0, -1\}$, which behaves as a smoothed differentiation and has generally shown good performance in comparison to Mel Frequency Cepstral Coefficients (MFCC), especially in noisy environments [9].

A feature vector of dimension 16 is calculated for each frame of the speech signal. Dynamic features are appended to the feature vector, either first derivative of the static feature vector and acceleration are included, the so called delta and delta-delta features. In addition, the delta coefficient of the logarithm of the frame energy is also computed. Thus, finally a vector of 49 features is obtained.

To deal with the background noise which degrades the speech signal we have used a spectral subtraction technique (SS) [13]. This estimate is initialized using the first frames of the utterance, and then recursively updated in those frames marked as non speech by a Voice Activity Detection (VAD) algorithm.

3.1.3. Back-end

The RAMSES recognizer is based on semi-continuous HMMs. We use 815 demiphone models [11] with 2 states per model and using a size of codebooks 512 (64 for the differential energy). Only those demiphone which appears in the training material a minimum of one hundred times are modeled, a total of 749. Those demiphones which frequency appearance is lower, the number of realizations is not enough to obtain a good HMM model, are modeled through a generalized demiphone model independently of context.

The language models were trained on text of the same topic used to build the geographical application corpus of ALBAYZIN. They are X-gram models which encode the syntax, semantics and pragmatics by concentrating on the local dependencies between words.

Four non-speech events: silence, speaker noise, filled pause and stationary noise are modeled each one by a four-state HMM. When decoding, they are combined in parallel, forming four possible alternatives, each with same probability.

3.2. GeoTALP-QA system

GeoTALP-QA has been developed within the framework of ALIADO[4] project. The system architecture uses a common schema with three phases that are performed sequentially without feedback: Question Processing (QP), Passage Retrieval (PR) and Answer Extraction (AE). More details about this architecture can be found in [5] and [14].

Before describing these phases, we introduce some additional geographical knowledge that have been used by our system for dealing with the geographic domain and some language-dependent NLP tools for Spanish.

3.2.1. Geographical Knowledge Base

A Geographical Knowledge Base (KB) of Spanish Geography has been built using the following resources: i) a set of $32,222$ non-ambiguous place names of Spain

[4] **ALIADO**. `http://gps-tsc.upc.es/veu/aliado`

with its geographical subclasses extracted from GEOnet Names Server (GNS[5]), ii) a set of 758 place names with its geographical subclasses from ALBAYZIN, iii) A set of patterns to create geographical NEs aliases (e.g. <toponym>_Mountains) iv) a lexicon containing 462 trigger words (e.g. “poblado" (ville)), v) a set of 7,632 groups of place names that refer to the same location [5], vi) a set of the most common trigger phrases (e.g. “city of *PLACE*", “*PLACE* Airport", etc.) in the geographical domain extracted from GNS gazetteer, vii) geographical subclasses, and viii) sets of words semantically related with each geographical subclass (e.g. "water" related with *sea* and *river*). The Geographical KB is used in an early stage of the QP phase to identify the geographical NEs of the transcribed question. Then the transcribed question is rewrited with the NEs in capital letters to allow that the NERC system could work properly. This KB is also used to obtain the possible geographical subclasses of each place name detected as NE.

3.2.2. Language-Dependent Processing Tools

A set of general purpose NLP tools are used for Spanish (see [15] for a more detailed description of these tools). *FreeLing*[6] is used for tokenization, morphological analysis, POS tagging, lemmatization, and partial parsing. The *ABIONET* NE Recogniser and Classifier (NERC) is used to tag NEs with basic categories (person, organization, location, and others). Finally, EuroWordNet is used to obtain a list of synsets, a list of hypernyms of each synset, and the Top Concept Ontology classes for each token. The same tools are used for the linguistic processing of both the questions and the passages.

3.2.3. Question Processing

The main goal of this subsystem is to detect the Question Type (QT), the Expected Answer Type (EAT), and the question analysis. We represent the lexical, syntactic and semantic information of the question and the passages. A language independent formalism, the *Environment*, represents the semantic content. The *Environment* of a sentence or a question contains the semantic relations that hold between the different components identified in the text. These relations are organised into an ontology of about 100 semantic classes and 25 relations. See below the classes related with the geographical domain.

```
ENTITY
    ENTITY_PROPER_PLACE
        GEOLOGICAL_REGION
            ARCHIPELAGO
            ISLAND
            LAND_FORM
                MOUNTAIN
            SEA_FORM
                CAPE
                GULF
                SEA
            WATER_FORM
                RIVER
        POLITICAL_REGION
            CITY
            CONTINENT
            COUNTY
            COUNTRY
            STATE
    ENTITY_QUANTITY
      NUMERIC
    MAGNITUDE
          AREA
          LENGTH
          FLOW
          WEIGHT
```

3.2.4. Question Classification

The Question Classification subsystem uses a Prolog DCG Parser with manually built rules to detect a set of 10 Question Types (see Table 1). This parser uses the following features: word form, word position in the question, lemma and part-of-speech (POS). The parser uses external information such: introductory phrases for each Question Type (e.g. "which extension" is a phrase of the QT *What_area*) and data from the Geographical KB (geographical NE subclasses, trigger words for each Geographical subclass, and semantically related words of each subclass).

Question Type	Expected Answer Type
Count_objects	NUMBER
How_many_people	NUMBER
What_area	MEASURE_AREA
What_flow	MEASURE_FLOW
What_height	MEASURE_HEIGHT
What_length	MEASURE_LENGTH
Where_action	LOCATION_SUBCLASS
Where_location	LOCATION_SUBCLASS
Where_quality	LOCATION_SUBCLASS
Default_class	LOCATION

Table 1. QTs and Expected Answer Types

3.2.5. Passage Retrieval

The PR approach uses the search-engine Google to get snippets. We use a boolean retrieval schema that takes advantage of the phrase search option of Google and the Geographical KB to create the queries.

The algorithm used to build the queries tries to maximize the number of relevant sentences with only one query per question. First, some expansion methods described below can be applied over the keywords. The expansion methods used are: i) join trigger phrases (e.g. "isla Conejera" o "Sierra de los Pirineos"), ii) expand NEs with trigger words (e.g. "Conejera" is expanded to: ("isla Conejera" OR "Conejera")), iii) Noun Phrase expansion using the groups of place names that refer to the same location, iv) Question-based Expansion (i.e. appending keywords or expanding the query depending on the question type). Then, stop-words (including some

[5] **GNS**. http://earth-info.nga.mil/gns/html
[6] **Freeling**. http://www.lsi.upc.edu/~nlp/freeling

trigger words) are removed. Finally, only the Nouns and Verbs are extracted from the keywords list.

3.2.6. *Answer Extraction*

Answer Extraction can be performed using two different components: a Semantic Relaxing algorithm and Frequency Based one.

Depending on the QT, a subset of useful relations of the question *Environment* has to be selected in order to extract the answer. We define this set of relations as the *Semantic Constraints* (SC) that are supposed to be found in the *Environment* of the answer sentence. These relations are classified as mandatory (MC), or optional (OC). Then, a set of extraction rules are applied following an iterative approach. In the first iteration all the MC have to be satisfied by at least one of the candidate sentences. Then, the iteration proceeds until a threshold is reached by relaxing the MC. The relaxation process of the SC is performed by means of structural or semantic relaxation rules, using the semantic ontology. The extraction process consists on the application of a set of QT dependent extraction rules on the set of sentences that have satisfied the MC. In order to select the answer from the set of candidates, the following scores are computed and accumulated for each candidate sentence: i) the rule score (which uses factors such as the confidence of the rule used, the relevance of the OC satisfied in the matching, and the similarity between NEs occurring in the candidate sentence and the question), ii) the passage score, iii) a semantic score (see [15]) , iv) the extraction rule relaxation level score. The answer to the question is the candidate with the best global score.

Finally, a Frequency Based Extraction component has also been implemented. This extraction algorithm is quite simple. First, all snippets are pre-processed. Then, we make a ranked list of all the tokens satisfying the expected answer type of the question (see the score formulae for each candidate token in Eq. 1). Finally, the top-ranked token is extracted.

$$Score(tk_i) = \sum_{o \in Occurrence(tk_i)} \frac{1}{snippet\_rank(o)} \quad (1)$$

4. EXPERIMENTS AND EVALUATION

4.1. ASR Evaluation

A test set of 722 utterances of clean speech about the Spanish geography, composed by roughly $6,600$ words, has been used to evaluate the performance of the ASR system, see section 3.1.1. A total of $20,496$ utterances from both close-talk and lavalier microphones of SPEECON database have been used to enroll the SCHMM models of the acoustic units. A lexicon of $1,086$ words is derived from the test set and the Language Model is also obtained from the test utterances. Acoustic and Language Models information are represented together by a Finite State Machine (FSA). During the testing phase, a Viterbi algorithm is in charge to make the search of the most probable sequence of demiphones. Table 2 shows the performance achieved by the ASR using several configurations of the weights of the probability path computed by Viterbi. These weights normalize the dynamic range of both acoustic and language model probabilities. Positive values of the additive weight S penalize probability transitions, hence the probability of word omissions increases. Higher values of the multiplicative weight M imply to place more confidence in the Language Model. We have fix to -1 the S value to obtain the results shown in Table 2.

	RAMSES performance		
Prob. weigth M	% Sentence	% Word	% Accuracy
9	75.90	93.88	93.82
7	76.87	94.44	94.39
5	75.90	94.32	94.26
3	64.54	90.88	90.74

Table 2. RAMSES-ASR percentage performance over the corpus testing set of ALBAYZIN about the Spanish Geography. Second and third column shows the percentage of the correct recognised sentences and words respectively

4.2. Geographical QA Evaluation

The GeoTALP-QA system has been evaluated with a test set of 62 questions about the Spanish geography (see [5] for more details about this evaluation). These questions were extracted from the geographical sub-corpus of ALBAYZIN [16], which contains utterances of questions about the geography of Spain. In Table 3 are shown the results in accuracy of the GeoTALP-QA evaluation using two different Answer Extractors: Semantic Relaxing and Frequency-Based.

	GeoTALP-QA Accuracy	
Num. Snippets	Semantic Relaxing	Frequency-Based
10	0.1774 (11/62)	0.5645 (35/62)
20	0.2580 (16/62)	0.5967 (37/62)
50	0.3387 (21/62)	0.6290 (39/62)

Table 3. GeoTALP-QA results over a test set of 62 questions about the Spanish Geography

5. CONCLUSIONS AND FUTURE WORK

We have presented GeoVAQA, a Restricted Domain Spoken Question Answering system in the scope of the Spanish geography. The aim of this paper has been to present an application be able to cope with spoken questions about the Spanish geography domain. The system consists in a web-based application which captures and transmits spoken questions, and presents a concise answer and a set of relevant snippets. In the server side, speech recognizer RAMSES is carrying out the speech transcription which becomes the input of the GeoTALP-QA system. Finally, the transcribed question is processed by the QA system and a concise answer goes back to the text display of the web client.

Both ASR and QA systems have been evaluated separately. The ASR achieves an accuracy around 94% and 76% of correct sentences over a corpus of 722 utterances. Most of the sentences errors were produced by only one substitution or omission. Therefore, if this error occurs to a non critical word, such as a connecting word, it is possible the error does not propagate to the QA system. However, if the word error is produced on a trigger or an introductory question word, it will degrade the performance of the retrieval system. That leads, for instance, to possible feedback strategies allowing QA system interacts with the confidence of the ASR recognition.

The QA system achieved a maximum accuracy of a 62.9% of correct questions over a corpus of 62 questions. This accuracy was obtained using the Frequency-Based algorithm and a set of 50 snippets from Google. On the other hand, the Semantic Relaxing algorithm has offered a lower performance than the Frequency-Based one. The Semantic Relaxing algorithm has achieved a maximum accuracy of 33.87% using the same number of snippets.

As a further work we plan to do an evaluation of the whole system. This evaluation is required to study the ASR degradation process and the drop in accuracy of the global system (ASR and QA) with respect of the results with only QA. An improvement of the results of the QA system is required to achieve a good accuracy during the answering process. The major efforts during the QA phase must deal with the reliability of the Passage Retrieval results (it implies working with the NEs and trigger word ambiguities) and the design of fast and precise Answer Extraction algorithms.

Acknowledgements

This work has been partially supported by the European Commission (CHIL, IST-2004-506909) and the Spanish Research Dept. (ALIADO, TIC2002-04447-C02). Daniel Ferrés is supported by a UPC-Recerca grant from Universitat Politècnica de Catalunya (UPC). TALP Research Center is recognized as a Quality Research Group (2001 SGR 00254) by DURSI, the Research Department of the Catalan Government.

6. REFERENCES

[1] Sanda M. Harabagiu, Dan I. Moldovan, and Joe Picone, "Open-Domain Voice-Activated Question Answering.," in *COLING*, 2002.

[2] Edward Schofield and Zhiping Zheng, "A speech interface for open-domain question-answering," in *ACL*, 2003, pp. 177–180.

[3] Svetlana Stenchikova, Dilek Hakkani-Tür, and Gokhan Tur, "QASR: Spoken Question Answering Using Semantic Role Labeling.," in *ASRU*, 2005.

[4] Chiori Hori, Takaaki Hori, Hajime Tsukada, Hideki Isozaki, Yutaka Sasaki, and Eisaku Maeda, "Spoken interactive ODQA system: SPIQA," in *ACL*, 2003, pp. 153–156.

[5] Daniel Ferrés and Horacio Rodríguez, "Experiments Adapting an Open-Domain Question Answering System to the Geographical Domain Using Scope-Based Resources.," in *MLQA, EACL*, 2006, pp. 69–76.

[6] A. Bonafonte, J.B. Mariño, A. Nogueiras, and J.A. Rodríguez, "RAMSES: el sistema de reconocimiento del habla continua y gran vocabulario desarrollado por la UPC," in *VIII Jornadas de Telecom I+D*, 1998.

[7] D.J. Iskra, B. Grosskopf, K. Marasek, H. van den Heuvel, F. Diehl, and A. Kiessling, "SPEECON - Speech Databases for Consumer Devices: Database Specification and Validation," in *LREC*, 2002, pp. 329–333.

[8] A. Moreno, D. Poch, A. Bonafonte, E. Lleida, J. Llisterri, J.B. Mariño, and C. Nadeu, "ALBAYZIN Speech Database: Design of the Phonetic Corpus," in *Eurospeech'93*, 1993, vol. 1, pp. 175–178.

[9] C. Nadeu, D. Macho, and J. Hernando, "Time and Frequency Filtering of Filter-Bank Energies for Robust Speech Recognition," in *Speech Communication*, 2001, vol. 34, pp. 93–114.

[10] P. Pujol, C. Nadeu, D. Macho, and J. Padrell, "Speech Recognition Experiments with the SPEECON Database Using Several Robust Front-Ends," in *ICSLP*, 2004.

[11] J.B. Mariño, A. Nogueiras, P. Pachés, and A. Bonafonte, "The Demiphone: an Efficient Contextual Subword Unit for Continuous Speech Recognition," in *Speech Communication*, 2000, vol. 32, pp. 187–197.

[12] J.B. Mariño, P. Pachés, and A. Nogueiras, "The Demiphone versus the triphone in a decision-tree state-tying framework," in *Proceedings of the ICSLP*, 1998.

[13] S.F. Boll, "A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech," in *ICASSP*, 1979, pp. 200–203.

[14] D. Ferrés, S. Kanaan, A. Ageno, E. González, H. Rodríguez, M. Surdeanu, and J. Turmo, "Structural and Hierarchical Relaxing of Semantic Constraints.," in *CLEF*, 2004, pp. 557–568.

[15] D. Ferrés, S. Kanaan, E. González, A. Ageno, H. Rodríguez, M. Surdeanu, and J. Turmo, "Structural and Hierarchical Relaxation Over Semantic Constraints," in *TREC*, 2004.

[16] J. Diaz, A. Rubio, A. Peinado, E. Segarra, N. Prieto, and F. Casacuberta, "Development of Task-Oriented Spanish Speech Corpora," in *ICLRE*, 1998, pp. 497–501.