Procesamiento del Lenguaje Natural, Revista nº 47 septiembre de 2011, pp 349-350 recibido 04-05-2011 aceptado 24-05-2011

Spoken language recognition in conversational telephone speech and TV broadcast news (GLOSA) *

Verificación de la lengua en conversaciones telefónicas y en informativos de televisión (GLOSA)

Luis Javier Rodríguez-Fuentes, Amparo Varona, Mikel Peñagarikano, Mireia Díez, Germán Bordel

Resumen: En esta breve comunicación presentamos el proyecto *GLOSA*, financiado por el Gobierno Vasco durante el bienio 2010-2011. El proyecto plantea, entre otros, los siguientes objetivos tecnológicos: (1) crear una infraestructura adecuada para desarrollar y evaluar nuevos métodos de verificación de la lengua; y (2) preparar un sistema competitivo de verificación de la lengua para señales telefónicas con objeto de presentarlo a la NIST 2011 Language Recognition Evaluation. Desde el punto de vista académico, el objetivo más importante del proyecto es la implementación y mejora de las técnicas actuales de verificación de la lengua.
Palabras clave: Verificación de la lengua, Indexado y búsqueda de audio/video

Abstract: In this brief communication we present the project *GLOSA*, financed by the Government of the Basque Country for the period 2010-2011. The project has two main technological objectives: (1) creating a suitable infrastructure for the development and evaluation of language recognition technologies; and (2) preparing a competitive language recognition system for conversational telephone speech, which will be eventually presented to the NIST 2011 Language Recognition Evaluation. From an academic point of view, the project aims to implement and improve state-of-the-art technology.
Keywords: Spoken Language Recognition, Audio/Video Indexing and Search

1. General description

Spoken language recognition may be required by applications dealing with multilingual speech. One of such applications is multimedia (audio/video) indexing from automatic speech transcriptions. In the last years, the Software Technologies Working Group (GTTS) of the University of the Basque Country has developed a prototype system for indexing and searching multimedia contents (*Hearch*, http://gtts.ehu.es/Hearch/). Currently, the system is able to transcribe and index speech from TV broadcast news in Basque and Spanish. A language recognition module can be easily integrated into the system to decide, for each segment, if it contains one of the target languages or some unknown language, either to apply a suitable set of acoustic and syntactic models in the Automatic Speech Recognition (ASR) module, or to discard the segment, which would not be transcribed and therefore would not produce index entries.

Our research group entered the field of language recognition in 2007, by presenting a fast but low-performance system to the NIST 2007 LRE. For the core task of the NIST 2009 LRE, which included conversational telephone speech in 23 target languages (including a sizeable amount of dialects) and many unknown languages, GTTS developed a quite competitive language recognition system. This project is partly motivated by the aim to improve and adapt that system to 16 kHz speech recorded from broadcast news, taking Basque, Spanish and English as target languages. By January 2010, we had developed a baseline GMM-UBM system for 16 kHz speech signals, but it didn't include English as target language and its performance was worse than the system developed for the NIST 2009 LRE.

* This project has been supported by the Government of the Basque Country under program SAIOTEK (project S-PE10UN87) and by the University of the Basque Country under grant GIU10/18.

ISSN 1135-5948

2. Objectives

2.1. Technological objectives

1. Improving the infrastructure for the development and evaluation of language recognition technologies. This involves increasing the computational and data storage resources currently available in the research group and creating broadcast news databases.
2. Developing a competitive language recognition system for conversational telephone speech, which will be eventually presented to the NIST 2011 Language Recognition Evaluation.
3. Developing and evaluating a spoken language verification system for 16 kHz speech, taking Basque, Spanish and English as target languages. This system will be eventually integrated in the backend of our audio/video search tool.

2.2. Scientific objectives

1. Analyzing state-of-the-art technology and implementing the best performing and/or the most feasible approaches, according to the available resources. As far as possible, open software and publicly available databases will be used, for the results to be more easily verifiable by other researchers.
2. Making scientific contributions for the improvement of previously developed baseline systems. Research will focus on two modeling approaches: phonotactic systems based on SVM classifiers and acoustic systems based on GMM and Joint Factor Analysis.

3. Project achievements and current state

Group infrastructure has been enhanced with the purchase of a computation server and several storage servers, partially financed by this project. The KALAKA database (Rodriguez-Fuentes et al., 2010a), created for the 2008 Albayzin LRE, has been extended with additional broadcast news recordings, leading to KALAKA-2, which involves 6 target languages (Basque, Catalan, English, Galician, Portuguese and Spanish) and includes around 125 hours of 16 kHz speech. KALAKA-2 has supported the 2010 Albayzin LRE (Rodriguez-Fuentes et al., 2010b), and provides the necessary datasets to develop a language recognition module for the backend of *Hearch*.

Research activitiy has been planned according to the available infrastructure and performance expectations based on our own preliminary experimentations and results reported by other authors. A competitive phonotactic system has been developed, based on publicly available phone decoders by the Brno University of Technology, using lattice-based phone n-gram counts, a simple but innovative feature selection algorithm which allows to effectively use high-order n-grams, and incorporating time-synchronous cross-decoder phone co-occurrences as high-level features (Penagarikano et al., 2011). This activity has led to a number of scientific contributions in the most relevant workshops, conferences and journals.

We are currently developing an acoustic system based on GMM and total factor analysis, which is expected to further improve the performance of the phonotactic system, by means of discriminative score fusion. The resulting (fused) system will be presented to the NIST 2011 LRE.

Bibliography

Penagarikano, M., A. Varona, L.J. Rodriguez-Fuentes, and G. Bordel. 2011. Improved Modeling of Cross-Decoder Phone Co-Occurrences in SVM-Based Phonotactic Language Recognition. *IEEE Transactions on Audio, Speech and Language Processing (Accepted)*.

Rodriguez-Fuentes, L.J., M. Penagarikano, G. Bordel, A. Varona, and M. Diez. 2010a. KALAKA: A TV Broadcast Speech Database for the Evaluation of Language Recognition Systems. En *7th International Conference on Language Resources and Evaluation (LREC 2010)*, Valleta, Malta.

Rodriguez-Fuentes, L.J., M. Penagarikano, A. Varona, M. Diez, and G. Bordel. 2010b. Overview of the Albayzin 2010 Language Recognition Evaluation: database design, evaluation plan and preliminary analysis of results. En *VI Jornadas en Tecnologias del Habla and II Iberian SLTech Workshop (FALA 2010)*, Vigo, Spain.