

RECONOCIMIENTO DE VOZ MULTIDIALECTAL ESPAÑA-COLOMBIA. TRABAJANDO CON FONEMAS Y SEMIFONEMAS

Mónica Caballero Galeote, Asunción Moreno Bilbao

Departamento de Teoría de la Señal y Comunicaciones
Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (TALP)
Universidad Politécnica de Cataluña
almcg@gps.tsc.upc.es, asuncion@gps.tsc.upc.es

ABSTRACT

It is well known that Spanish is one of the most spoken languages in the world, however, Spanish presents differences depending on the region where it is spoken. In this paper, we will see how these differences, and especially differences in pronunciation, affect in speech recognition.

This paper reflects the results of the continuation of a project on multi-dialectal speech recognition. More important conclusions about the main idea of this project, which is to use the information from the most common dialect to improve speech recognition systems of others dialects of the same language, can be taken now based on the new results that are reported in this paper.

1. INTRODUCCIÓN

El idioma español se encuentra entre los idiomas más hablados en todo el mundo. Se habla en España y en toda Latinoamérica, exceptuando Brasil por más de 300 millones de personas. También es verdad que debido a la gran dispersión geográfica de estas zonas, el español de cada zona difiere y se establecen variedades dialectales.

Es por eso, que al plantearse un sistema de reconocimiento del habla para el castellano, parece interesante no limitarse a un solo dialecto, sino intentar crear un sistema multidialectal.

Más aún si se observa que para el castellano canónico o estándar, el hablado en España, se tienen bases de datos de habla de 1000 y 4000 locutores, que son bastantes recursos. Pero hay muchos dialectos del español de Latinoamérica de las que sólo se tienen pequeñas bases de datos, como es el caso del dialecto hablado en Colombia [1].

Lo que se plantea en este proyecto es el poder aprovechar todos los conocimientos y datos que se puedan extraer de una gran base de datos para mejorar sistemas de reconocimiento del español hablado en otras zonas, con otras características fonéticas, con diferencias léxicas, pero a fin de cuentas, el mismo idioma.

Este estudio es una aproximación al sistema multidialectal que se plantea, dónde se trabaja con un solo dialecto latinoamericano, el hablado en Colombia además del castellano hablado en España.

2. SISTEMA DE RECONOCIMIENTO

2.1. Descripción del problema fonético

Se ha hecho un estudio de las reglas de transcripción grafema-fonema del español de Bogotá, la zona más poblada de Colombia, dónde se habla el dialecto propio de las montañas o andino, así como también del castellano estándar hablado en España [2]

Las transcripciones se han realizado en símbolos SAMPA (Speech Assessment Methods Phonetic Alphabet) y algunos símbolos de SAMPA extendido para representar sonidos típicos colombianos, como el caso de /h/.

Las reglas para la transcripción del colombiano se basan en la transcripción del castellano estándar. Así, respecto a éste caracterizan el dialecto colombiano los siguientes fenómenos:

- Pronunciación de /T/ siempre como /s/, el “seseo”.
- La /L/ siempre se pronuncia como /jj/, el “yeísmo”.
- La velar fricativa /x/ se realiza como la glotal /h/.
- Los sonidos /b/, /d/ y /g/ se pronuncian siempre como consonantes oclusivas excepto cuando se encuentran en posición post-nuclear o en el inicio de una sílaba siguiendo a una vocal. En estos casos se transcriben como los alófonos aproximantes /B/, /D/ y /G/.

Dadas estas características se toman estas decisiones a la hora de transcribir:

- Para el castellano: En castellano casi no se realiza la /N/. Por lo que se asimila la /N/ a la /n/. Igual pasa con los sonidos /b/, /d/ y /g/ que se asimilan a los alófonos aproximantes /B/, /D/ y /G/. También se considera que los sonidos /L/ y /jj/ están demasiado cercanos y se decide asimilarlos a uno común que se etiqueta como /y/.
- Para el colombiano: Se transcribe siempre la /T/, la /L/ y la /x/ como /s/, /jj/ y /h/. En cambio se diferencian /b/ /, /d/ y /g/ de /B/, /D/ y /G/. Y al igual que en el castellano se asimila la /N/ a la /n/.
- Para el entrenamiento conjunto: Se comprobó que para entrenar con las dos bases a la vez, es mejor separar los fonemas /B/, /D/ y /G/ de /b/, /d/ y /g/ en castellano. En colombiano ya se distinguen. En cuanto a la /L/ y la /jj/, se

mantienen asimiladas y se transcriben como /jj/, al igual que en colombiano.

2.2. Descripción general del sistema

En este estudio se ha trabajado con un sistema de reconocimiento desarrollado por el grupo de procesado del habla del centro de Tecnologías aplicadas al lenguaje y la habla (TALP) de la UPC llamado RAMSES [3]

Se parametriza con mel-cepstrum [4] y se crea un codebook de 128 coeficientes. Se entrenan modelos semicontinuos de Markov. Se ha trabajado con fonemas y semifonemas incontextuales y semifonemas por umbral dependientes del contexto.

3. EXPERIMENTOS Y RESULTADOS

Se diseñaron una serie de experimentos con el objetivo de observar no sólo los datos finales del reconocimiento, sino de ver qué factores afectaban la mejora de estas tasas.

3.1. Ficheros entrenamiento y test

Se pretende evaluar el sistema reconociendo palabras de aplicación y dígitos aislados. Para éste propósito se decide entrenar con frases fonéticamente balanceadas y ampliar el entrenamiento con palabras fonéticamente balanceadas.

- *Entrenamiento.* En Colombia se tienen 2700 frases pertenecientes a 800 locutores. En castellano, se dispone de un total de 15156 frases y 9174 palabras pertenecientes a 3500 locutores.
- *Test.* Se diseñan dos tests, uno para el español y otro para el colombiano con las palabras de aplicación y dígitos aislados. En colombiano se tienen un total de 751 ficheros de test de 200 locutores. En Castellano se dispone de 626 palabras pertenecientes también a 200 locutores.

En total se entrenan un total de 26 fonemas en castellano y de 30 en colombiano y se reconocen un total de 30 palabras de aplicación más los 10 dígitos en los dos casos.

3.2. Experimentos y resultados

La primera parte de este estudio consistió en una aproximación trabajando con fonemas[5]. Se evaluó cada sistema por separado y también se diseñaron una serie de experimentos de reconocimiento cruzado. A la vez que se iban obteniendo modelos nuevos del castellano, se les aplicaba el test colombiano. Los resultados finales a los que se llegó son los siguientes:

ENTRENAMIENTO(*)	TEST	RESULTADO
800S Colombia	Colombiano	94,14 %
3500SW España	Castellano	94,40 %
3500SW España	Colombiano	94,27 %

(*)ENTRENAMIENTO: N° Locutores; S= Sentence (frase), W=Word (palabra); Dialecto

Con los mejores resultados se entrenaron modelos de semifonemas contextuales por umbral, utilizando como modelos iniciales semifonemas incontextuales.

ENTRENAMIENTO	TEST	RESULTADO
800S Colombia	Colombiano	97,86 %
3500SW España	Castellano	97,60 %
3500SW España	Colombiano	97,33 %

Por otra parte también se han hecho experimentos con entrenamiento conjunto, entrenando con las 2 bases a la vez. Los resultados obtenidos hasta el momento trabajando con fonemas son los siguientes:

ENTRENAMIENTO	TEST	RESULTADO
800S CO + 800SW ES	Colombiano	93,87 %
800S CO + 800SW ES	Castellano	94,40 %

4. CONCLUSIONES

Con el estudio en fonemas se observaba que era posible conseguir unas buenas tasas de reconocimiento utilizando los modelos entrenados con la base de datos española. El paso a semifonemas nos hace ganar de media tres puntos en las tasas de reconocimiento, a la vez que nos muestra uno de los mayores defectos del dialecto castellano. Éste es su gran variabilidad. El hecho de utilizar semifonemas contextuales le va mejor a Colombia, sin embargo España no aumenta tanto, a pesar de contar con más locutores. Esto también se ve reflejado en el reconocimiento cruzado, ya que los semifonemas castellanos ya no modelan tan bien a los colombianos, como pasaba con los fonemas.

En cuanto al entrenamiento conjunto se observa que ninguna de las dos bases hace que las tasas empeoren significativamente, sino que se complementan. Es de esperar, y es la línea de trabajo que queda marcada, que al añadir locutores españoles, y contando con las combinaciones de semifonemas propias del colombiano, la tasa se incremente.

5. AGRADECIMIENTOS

Este trabajo ha sido subvencionado por el proyecto TIC-2000-1005-C03.

6. REFERENCIAS

- [1] A. Moreno, H. Höge, J. Koehler, J. B. Mariño. *Project SALA, SpeechDat Across Latin America.*, LREC'98, Granada.
- [2] A. Moreno, J. B. Mariño. *Speech Dialects: Phonetic Transcription.* ICSLP, Australia. Dec, 1998.
- [3] A. Bonafonte, J. B. Mariño, A. Nogueiras, J.A. Rodríguez Fonollosa. *RAMSES: el sistema de reconocimiento del habla continua y gran vocabulario desarrollado por la UPC.* (TELECOM I+D'98), Madrid.
- [4] J.W. Picone *Signal Modeling Techniques in Speech Recognition.* Proceedings of the IEEE Vol 81 no9. September 1993
- [5] M. Caballero, A. Moreno. *Reconocimiento de voz multidialectal; España-Colombia.* Buran, nº17, 2001.