



POLITÉCNICA

Año Académico 2008/2009
Plan 94

**UNIVERSIDAD POLITÉCNICA DE MADRID
E.T.S. DE INGENIEROS DE TELECOMUNICACION**

ACTA DE EXAMEN

Asignatura: PROYECTO FIN DE CARRERA

TÍTULO DEL PROYECTO: DESARROLLO, IMPLEMENTACIÓN Y PRUEBAS DE TÉCNICAS DE INDEXADO DE LOCUTORES EN DOCUMENTOS HABLADOS

APELLIDOS Y NOMBRE: FERNÁNDEZ GARCÍA, CÉSAR PABLO	CALIFICACIÓN <i>10,00 MATRÍCULA 1 DE HONOR</i>
---	---

Tutor: D. JAVIER FERREIROS LOPEZ

Madrid, a *25 de Septiembre de 2009*

EL VOCAL PRIMERO

EL PRESIDENTE

EL VOCAL SECRETARIO

MIEMBROS DEL TRIBUNAL

Presidente: D. JOSE MANUEL PARDO MUÑOZ

Vocal: D. JAVIER FERREIROS LOPEZ

Secretario: D. RUBEN SAN SEGUNDO HERNANDEZ

Resumen

El indexado de locutores (*Speaker diarization*) es una tarea aplicable sobre cualquier documento hablado o grabación, y su principal objetivo es fragmentar la señal acústica en segmentos de voz que se asignan a diferentes locutores. Se trata, por tanto, de dar respuesta a la pregunta “¿quién ha hablado, y cuándo?”. En la modalidad seguida en el presente proyecto, esta tarea se realiza sin conocimiento previo sobre la cantidad ni la identidad de los hablantes.

La base de datos con la que se ha trabajado es un conjunto de 12 grabaciones de desarrollo correspondientes a diferentes reuniones (*Devel06*) que son un compendio de los datos entregados por el *National Institute of Standards and Technology (NIST)*. De este mismo organismo proceden los criterios y las métricas de evaluación con las que se examinan los resultados del sistema implementado de indexado de locutores, cuyo principal exponente es la *Diarization Error Rate (DER)*, que da idea del error que se ha cometido.

El sistema implementado parte de una parametrización del audio muestreado empleando coeficientes cepstrales en la escala Mel (*MFCC*). El bloque principal del sistema lo forman un algoritmo de detección de cambio acústico (*ACD*), basado en el criterio *BIC (Bayesian Information Criterion)*, que delimita segmentos de audio en los que sólo hay un hablante interviniendo en la conversación; y un algoritmo de agrupamiento o *clustering* que clasifica todos los segmentos generados por el *ACD* según el hablante del que proceden. La información de *clustering* se realimenta al *ACD* para disponer de más información con la que el *BIC* pueda decidir.

Como paso previo al *ACD* se han implementado diversos métodos sencillos de segmentación preliminar y descarte de silencios, basados en distancias y en energías de la señal. El objetivo de estos métodos es reducir la cantidad de información que tiene que procesar el *ACD*, generando una serie de candidatos a puntos de cambio acústico. Esto debe realizarse sin que se introduzca un error que se arrastre a lo largo del resto del sistema.

Se han obtenido y analizado una gran cantidad de datos de evaluación, probando y ajustando los diversos parámetros que intervienen en los distintos algoritmos del sistema, y se han presentado las conclusiones sobre de qué manera se pueden obtener las mejores prestaciones.

Palabras clave

Indexado de locutores, diarization, ACD, clustering, BIC, reuniones, audio, MFCC, distancias, grabaciones, segmentación, DER, Miss, False Alarm.

PROYECTO FIN DE CARRERA

**Desarrollo, implementación y pruebas
de técnicas de indexado de locutores en
documentos hablados.**

Alumno: **César Pablo Fernández García**
Tutor: **Javier Ferreiros López**
Departamento: **Ingeniería Electrónica**

Tribunal:

Presidente: **José Manuel Pardo Muñoz**
Secretario: **Javier Ferreiros López**
Vocal: **Rubén San Segundo Hernández**
Vocal suplente: **Juan Manuel Montero Martínez**

Fecha de lectura y defensa:

25 de Septiembre de 2009

Calificación:

M. G. P. MATRÍCULA DE HONOR