PLAN DE EVALUACIÓN PARA SEGMENTACIÓN E IDENTIFICACIÓN DE LOCUTORES

1.- Introducción

El presente Plan de Evaluación para Segmentación e Identificación de Hablantes proporciona las bases para la evaluación de aplicaciones en este campo de Tecnologías del Habla.

El objetivo de esta evaluación es fomentar los trabajos de investigación en relación con la segmentación de la voz en conversaciones con varios locutores, y además, la identificación parcial de algunos de éstos. Con esta finalidad se plantea un plan de evaluación que consiste en la segmentación e identificación de las intervenciones de hasta 5 locutores conocidos, en archivos donde podrán existir intervenciones de otros locutores. Ambos aspectos, segmentación e identificación, son importantes: una buena segmentación pero una mala identificación será considerada como una segmentación errónea. Durante la fase de entrenamiento o desarrollo se ofrecerán grabaciones cortas correspondientes a los 5 locutores a segmentar e identificar.

Los participantes se comprometen a presentar los resultados de la evaluación en una sesión especial que tendrá lugar durante las IV Jornadas en Tecnología del Habla. La participación se realiza a modo individual o en equipo formado por un máximo de 3 investigadores donde al menos uno de ellos debe ser estudiante.

2.- Medición de prestaciones

Para la evaluación se utilizarán la herramienta del NIST “md-eval-v21.pl” (The diarization evaluation tool) disponible en http://www.nist.gov/speech/tests/rt/rt2006/spring/[1]

El modo de funcionamiento de esta herramienta se ha tomado como referencia para la definición del formato de los archivos de resultados (formato RTTM) que se presenta en el punto 4.

3.- Condiciones de evaluación

El material de entrenamiento consistirá en 5 archivos de entre 1 y 5 segundos, cada uno de ellos con la voz de un locutor.

Todos los archivos contendrán muestras digitalizadas de audio con el siguiente formato:

[1] Gracias a Xavier Anguera por las gestiones para disponer de esta herramienta de evaluación

- Frecuencia de muestreo: 16 KHz
- Canales de audio: mono
- Tamaño de muestra: 16 bits
- Alineamiento de octeto: LSB primero (little endian).
- Ficheros PCM sin cabecera ni compresión.

Los archivos de audio tendrán la extensión **.pcm**.

4.- Formato de los archivos de etiquetas a generar

Los resultados de la segmentación se deberán presentar en archivos de texto (hipótesis), con el mismo nombre correspondiente al archivo de audio de partida, pero con extensión **.hyp**, y cuyas líneas contendrán la información temporal del punto de comienzo y duración de cada locutor, y una mínima identificación de éste. Un ejemplo, para un archivo *fichero21.pcm*, es:

Archivo: *fichero21.hyp*

```
SPEAKER FICH21 1 0.00 257.93 <NA> <NA> OTROS <NA>
SPEAKER FICH21 1 257.93 12.15 <NA> <NA> LOC2 <NA>
SPEAKER FICH21 1 270.40 25.60 <NA> <NA> OTROS <NA>
...
```

Cada intervención de un locutor se muestra en una línea diferente que contiene los siguientes campos:

- Modo de obtención de la señal: en todos los casos consideraremos este campo igual a SPEAKER.
- Número del fichero a considerar: FICH1, FICH2, ... hasta FICH20
- Canal: en todos los casos consideraremos el cana 1.
- Inicio: representa el segundo y centésima de segundo de comienzo del locutor.
- Duración: representa el segundo y centésima de segundo de la duración de la intervención del locutor.
- Campo 6: <NA>
- Campo 7: <NA>
- Código del locutor. Los códigos posibles para identificar al locutor con los siguientes: LOC1, LOC2, LOC3, LOC4, LOC5 y OTROS. La etiqueta OTROS corresponde con cualquier otro locutor/música/... que no corresponda a alguno de los 5 locutores conocidos.
- Campo 9: <NA>

La representación numérica del tiempo será en segundos y centésimas, pudiendo omitirse los ceros iniciales y finales, así como el punto decimal según los convenios habituales. El carácter decimal deberá ser el '**.**'. Las cantidades deberán estar separadas por espacios.

Estas hipótesis se compararán mediante la herramienta mencionada con una segmentación de referencia, contenida en archivos con el mismo formato y nombrados con la extensión **.ref**.
Un aspecto que conviene remarcar es no se han etiquetado las pausas breves (menores de 500 ms) entre dos intervenciones de un mismo locutor, considerándose como un mismo fragmento asociado a dicho locutor.

5.- Datos de evaluación

El material de prueba consistirá en 20 ficheros de entre 3 y 5 minutos con intervenciones de estos 5 locutores y/o de otros locutores.

El formato de estos archivos de audio será el mismo que el de los archivos de entrenamiento y se nombrarán con la misma extensión (**.pcm**).

6.- Procedimiento para la evaluación

El procedimiento con las fechas para la evaluación es el siguiente:

- El 21 de Julio de 2006 se dispondrá de los planes de evaluación y se abre el periodo de inscripción.
- La fecha límite de inscripción será el 15 de Septiembre de 2006.
- A partir del 16 de Agosto de 2006 se podrá disponer del material de entrenamiento y desarrollo para las distintas evaluaciones. Es necesario estar inscrito en la evaluación para recibir el material.
- El 18 de Octubre de 2006 se liberarán las bases de datos para la evaluación.
- El 27 de Octubre de 2006 a las 24:00 es la fecha límite para recibir los resultados en el formato y método indicados.
- El 3 de Noviembre de 2006 se enviarán los resultados de la evaluación.

7.- Envío de segmentaciones

Las segmentaciones se enviarán por correo electrónico a la organización. Deberán ser completas, conteniendo por tanto todo el conjunto de datos de evaluación.

Los ficheros de texto con las segmentaciones, con la extensión "**.hyp**", deberán enviarse comprimidas en un archivo ***.zip** a: **Rubén San-Segundo Hernández** (lapiz@die.upm.es)

Los resultados estarán disponibles una vez se hayan enviado dichos resultados a los participantes. Esto permitirá realizar análisis previos a la celebración de las IV Jornadas de Tecnologías del Habla.

Cada participante deberá remitir una descripción del sistema enviado a la evaluación, que deberá incluir:

- Nombre del sistema
- Condiciones de evaluación (base de datos de entrenamiento)
- Descripción de la aproximación algorítmica

Esta descripción se enviará en formato de texto ASCII o PDF. Las descripciones recibidas se distribuirán como parte del material de análisis de la evaluación.