

Apéndice B. Experimentos independientes del locutor en el sistema de reconocimiento de nombres deletreados.

Si analizamos en detalle la descripción de la base de datos utilizada para el desarrollo del reconocedor de nombres deletreados (apartado 3.6), podemos ver que los 1.000 ficheros que contienen las secuencias de letras aleatorias (una secuencia para cada uno de los 1.000 locutores), forman parte siempre del conjunto de entrenamiento. En consecuencia, para cada uno de los 6 entrenamientos realizados, un fichero de los locutores utilizados en validación y test (y sus condiciones de entorno correspondientes) está presente en el conjunto de entrenamiento. Para analizar su impacto sobre las tasas, hemos repetido los experimentos correspondientes a las secciones 3.7.1, 3.7.2 y 3.7.3 eliminando los ficheros con las cadenas de letras aleatorias de los locutores utilizados tanto en evaluación como en test (300 ficheros). Estos experimentos incluyen el sistema base, la incorporación de modelos de silencio contextuales y la consideración de modelos de ruido. Los resultados se presentan en la tabla B-1.

Sistema	Algoritmo One-pass					Etapa de hipótesis	
	Sus (%)	Ins (%)	Bor (%)	TEL (%)	PCP (%)	TE (%)	TP (xRT)
Sistema de Referencia	20,8	6,7	3,8	31,3	20,8	16,9	0,9
Silencios Contextuales	17,7	4,4	2,4	24,5	27,2	11,4	1,2
Silencios contextuales + Modelos de ruidos	16,0	1,2	2,5	19,7	33,8	8,3	1,2

Tabla B-1: Resultados del Sistema de Referencia, nueva topología de HMM con silencios contextuales y considerando los modelos de ruido. Se han eliminado 300 ficheros de cadenas aleatorias correspondientes a los locutores de validación y test: porcentajes de Sustituciones (Sus), Inserciones (Ins), Borrados (Bor), Tasa de Error de Letra (TEL), Porcentaje de Cadena Perfectas (PCP), y la Tasa de Error (TE) y el Tiempo de Proceso (TP) de la etapa de hipótesis.

En este caso, la incorporación de modelos de silencio contextuales en el propio modelo de letra ha reducido 6,8 puntos el error de letra (de 31,3% a 24,5%) consiguiendo un error de nombre de 11,4% en la etapa de hipótesis.

Como podemos ver, al eliminar los 300 ficheros las Tasas de Error (en comparación con los resultados del apartado 3.7.3) han aumentado ligeramente. Este incremento se ha producido principalmente por la reducción del número de ficheros utilizados para entrenar los modelos acústicos (1.800 en lugar de 2.100) y no por quitar los ficheros correspondientes a los locutores de validación y test. Para comprobar esta afirmación hemos realizado los siguientes experimentos utilizando en todos los casos los silencios contextuales y sin utilizar los modelos de ruido. Considerando un único Round Robin (1-Round Robin) hemos dividido el conjunto de test en 3 subconjuntos y hemos

entrenado modelos acústicos independientes para reconocer cada uno de los subconjuntos. Eliminando los datos de los locutores de validación y test, disponemos de 2000 ficheros para entrenar cada grupo de modelos acústicos. Los resultados finales son la media de los obtenidos en los 3 subconjuntos. De la misma manera hemos dividido los conjuntos de validación y test en 10 subconjuntos diferentes, considerando en cada caso 2.070 ficheros para entrenar los modelos acústicos. Los resultados se muestran en la tabla B-2.

Experimentos	TEL (%)	TEN (%)
Punto de referencia (2.100)	24,4	11,5
Eliminando 300 ficheros de locutores de validación y test (1.800)	25,2	12,5
Eliminando 100 ficheros de locutores de validación y test (2.000)	24,5	11,8
Eliminando 30 ficheros de locutores de validación y test (2.070)	24,4	11,6

Tabla B-2: Resultados con la topología HMM con silencios contextuales, eliminado 300, 100 y 30 ficheros respectivamente del conjunto de entrenamiento (considerando únicamente 1-Round Robin)¹.

Como podemos ver cuando se reduce el número de ficheros que se eliminan del conjunto de entrenamiento, los resultados mejoran alcanzando casi los mismos valores que los obtenidos en el caso de utilizar todos los ficheros para entrenar los modelos acústicos. En este último caso, cuando se prueba con un fichero de un locutor, en el conjunto de entrenamiento hay sólo un fichero (de entre 2.100 ficheros) que pertenece al mismo locutor, luego su influencia es irrelevante.

En este apéndice no hemos repetido los experimentos para las situaciones en las que se consideran los modelos de lenguaje 2-gram o 3-gram porque en estos casos, el decremento en la calidad del modelado acústico es menos significativo.

¹ Notar que los resultados presentados en esta tabla son diferentes de los presentados en las tablas 3-6 y 3-7 puesto que los datos utilizados son diferentes: en las tablas 3-6 y 3-7 se utilizó un 6-Round Robin mientras que en este caso se ha utilizado únicamente un 1-Round Robin.