VI Congreso Iberoamericano de Acústica - FIA 2008
Buenos Aires, 5, 6 y 7 de noviembre de 2008

FIA2008-A074

Impacto de la segmentación automática de voz en el entrenamiento de Modelos Ocultos de Markov con corpus diferentes

Mónica Godoy Millán[(a)],
Maria E. Vaca Navia[(b)],
Rafael A. Jordán Osorio[(c)],
Dr. Diego L. Linares Ospina [(d)]

(a) Estudiante de Ingeniería de Sistemas y Computación. Miembro del grupo de Investigación DESTINO. Pontificia Universidad Javeriana - Cali. Calle 18 No. 118-250. Santiago de Cali, Colombia. E-mail: mgodoy@javerianacali.edu.co

(b) Estudiante de Ingeniería de Sistemas y Computación. Miembro del grupo de Investigación DESTINO. Pontificia Universidad Javeriana - Cali. Calle 18 No. 118-250. Santiago de Cali, Colombia. E-mail: mevaca@javerianacali.edu.co

(c) Profesor Asistente del Departamento de Ciencias e Ingeniería de la Computación, área de Teoría y Modelos Computacionales. Miembro del grupo de Investigación DESTINO. Pontificia Universidad Javeriana - Cali. Calle 18 No. 118-250. Santiago de Cali, Colombia. E-mail: rjordan@javerianacali.edu.co

(d) Coordinador Institucional de Investigación. Profesor Asociado del Departamento de Ciencias e Ingeniería de la Computación, área de Teoría y Modelos Computacionales. Miembro del grupo de Investigación DESTINO. Pontificia Universidad Javeriana - Cali. Calle 18 No. 118-250. Santiago de Cali, Colombia. E-mail: dlinares@javerianacali.edu.co

Abstract

In order to train the Hidden Markov Models (HMMs), it is required to have a corpus whose waves have been segmented by one person. However, the particular way that each person has to segment means that HMMs have different characteristics, making phoneme's characteristics dependant on the person who segmented the corpus. This paper compares the segmentation made by two different people, how each corpus affects the HMMs training, and the boundaries generated by these training. We evaluate the efficiency of Automatic Segmentation Vs. Handle Segmentation. Finally, we set out the results and conclusions of this job.

Resumen

Para entrenar los Modelos Ocultos de Markov (HMM: Hidden Markov Models) fue necesario contar con un conjunto de ondas de voz o emisiones que se definen como corpus, estas emisiones por lo general son segmentadas por una persona, sin embargo la forma particular que tiene cada persona para segmentar hace que los HMMs tengan características diferentes, haciendo que éstas dependan de quien segmentó el corpus. En este trabajo se comparó la segmentación manual realizada por dos personas, cómo cada corpus afecta al entrenamiento de los HMMs y las fronteras de los fonemas que fueron generadas por estos entrenamientos. Finalmente, se exponen los resultados obtenidos y las conclusiones de este experimento.

1 Introducción

Un aspecto fundamental en el reconocimiento del habla es el de utilizar grandes corpus con su representación fonética y las fronteras de ésta, a esto se le llamará etiquetar y segmentar, respectivamente. Un corpus se puede etiquetar en unidades fonológicas llamadas fonemas, o conjunto de fonemas llamados palabras o de un conjunto de palabras llamados frases, esta forma de etiquetar dependerá de la clase de experimento que se quiera realizar o de los requerimientos. Es muy común que las etiquetas se realicen por fonema, pues estos son considerados la unidad fonológica más pequeña en que puede dividirse un conjunto fónico [1], y de esta forma se tiene un modelo de cada uno por separado, pudiendo así identificarlos sin tener un corpus por cada palabra o frase, ya que los fonemas son menos que las palabras o frases. Hoy en día la segmentación más precisa se realiza manualmente sin embargo este no es un trabajo sencillo, ya que resulta muy extenuante, costoso, demanda mucho tiempo para ser completado, exige experiencia lingüística, los errores humanos asociados se incrementan y además tiene el inconveniente de la discrepancia de criterios si la realizan personas diferentes.[2] Por tal motivo partimos de la hipótesis de que la segmentación manual realizada por dos o más personas produce grandes cambios en la segmentación automática generada por dicha segmentación. Estos cambios se refieren a la duración de los fonemas en la emisión y al inicio y final de las fronteras de los fonemas.

El objetivo de la segmentación automática consiste en generar un conjunto de segmentos que delimiten todas las unidades fonológicas a partir de emisiones y de la trascripción fonética correspondiente. Existen dos modelos para la segmentación automática, el más empleado es el análisis estocástico el cual tiene distintas técnicas que se basan en la teoría de la decisión de Bayes, la teoría de la información, las técnicas de comparación de patrones utilizando programación dinámica y modelos ocultos de Markov. El otro modelo son los no-estocásticos que se representan con Redes Neuronales, actualmente se han conseguido resultados comparables a los obtenidos con los HMM. Los dos modelos presentan diferentes problemas o inconvenientes de implementación, memoria y tiempo. [3]

Este documento muestra un experimento acerca del entrenamiento de HMMs, específicamente de cómo el corpus de entrenamiento puede afectar la segmentación del corpus de prueba y posteriormente el de evaluación. El experimento fue realizado bajo la herramienta HTK, la cual fue desarrollada en la Universidad de Cambridge. El documento empieza con una pequeña ilustración de los modelos y herramientas usadas, después se verá la explicación del método, la definición de los corpus empleados, la estimación de los parámetros que han sido ajustados para el español caleño hablado en Colombia en el departamento del Valle del Cauca, como se obtuvieron los corpus y los detalles de la segmentación fonética. Luego se pasa al desarrollo del experimento, en el que se muestra el diseño implementado, específicamente la topología del HMM que se usó, finalmente se muestran los resultados y conclusiones.

2 Desarrollo del artículo

Para la segmentación automática o manual es necesario contar con un corpus inicial porque con éste se entrenarán las emisiones o se compararán los resultados obtenidos. Las características del corpus dependerán del tipo de experimento a realizar y sus objetivos, para nuestro experimento se utilizarán dos corpus que tienen condiciones diferentes, mientras uno fue grabado en un estudio de grabación con aislamiento de ruido, software y micrófonos especiales, el otro no. Como evidencia se muestran a continuación algunas aplicaciones y características de los corpus utilizados en la actualidad.

- El sistema SAPLEN [4]

El objetivo es evaluar sistemas conversacionales mediante la simulación de las interacciones que realizan los usuarios en la vida real.

Para construir el corpus de frases, se han seleccionado aleatoriamente 500 frases de un corpus de 523 diálogos previamente obtenidos en un restaurante de comida rápida. Las transcripciones fonéticas de las 500 frases se han creado manualmente, y las representaciones semánticas correspondientes se han creado automáticamente. Nueve locutores han grabado varias versiones de las 500 frases seleccionadas.

- Proyecto SenSem [5]

El objetivo es construir un banco de datos de verbos del español. Dicho banco reflejará el comportamiento sintáctico semántico de 250 verbos del español.

El corpus se compone de 25.000 oraciones del español anotadas a nivel sintáctico-semántico. Estos 25.000 ejemplos ilustran el comportamiento de los 250 verbos usados con más frecuencia en el español, según datos estadísticos extraídos de un corpus periodístico de más de 13 millones de palabras. Se han extraído de forma aleatoria 100 ejemplos correspondientes a cada verbo de un corpus de la versión electrónica de El Periódico de Catalunya. Los ejemplos excluyen los usos perifrásticos de los verbos, así como expresiones idiomáticas y colocaciones. En total, el corpus contiene 750,000 palabras, de las cuales 350,000 están dentro del alcance de algún verbo que ha sido anotado.

El corpus es una de las partes más importante en esta clase de experimentos, debe ser lo suficientemente confiable porque de éste dependerán los resultados y los avances del experimento. Este artículo consiste en comparar los resultados que se obtienen de dos corpus que han sido etiquetados por dos personas con conocimientos lingüísticos. Si se pueden comparar sus etiquetas y segmentos, si los lingüistas segmentan de la misma forma, además si la segmentación automática tiene grandes diferencias en comparación a la manual. Pero, ¿cómo comprobar que la segmentación automática de una emisión es “muy similar” a lo que haría un lingüista, si entre estos la forma de segmentar es distinta? Con esta incertidumbre y muchas otras, mostraremos nuestra experiencia en la segmentación del corpus y los resultados que el experimento arroja.

2.1 Explicación del método

Como se explicó anteriormente, el corpus constituye una parte fundamental en el reconocimiento del habla, pues a partir de las características de los fonemas incluidos en el corpus, se puede obtener un modelo para reconocer o segmentar nuevas emisiones. De esta manera, el corpus que se emplee en un experimento afectará sustancialmente los resultados que se obtengan. Por esta razón, es vital ser muy cuidadosos a la hora de definir el corpus (la gramática, los locutores, el estudio de grabación, etc.) pues los resultados, dependerán completamente del cuidado que se ponga en la obtención del mismo. Además, estos también dependen de la persona que lo haya segmentado, en el caso que el corpus haya sido segmentado manualmente. Esto sucede porque cada persona tiene su manera particular de segmentar, además de otros factores que pueden afectar como son: la habilidad para diferenciar un fonema de otro, la experiencia, la preparación académica, fisionomía, entre otros.

Al tener una forma particular de segmentar, las características fonéticas de los modelos no serán iguales, y los resultados que se obtengan probablemente serán diferentes. Para

demostrar que tan diferentes pueden ser los resultados, se realizó este experimento el cual pretende mostrar como la segmentación manual afecta las fronteras de nuevas emisiones segmentadas bajo estos corpus. Para esto, se requiere un sistema de segmentación automático confiable, cuya segmentación sea similar a como lo haría un lingüista.

A continuación, se hablará sobre los corpus usados en el experimento, como se obtuvieron, las características técnicas y demás detalles.

2.1.1 Obtención del Corpus

La generación del corpus es una de las tareas más importantes, pues es la clave para obtener los modelos ocultos de los fonemas que se entrenan con la herramienta HTK. En esta etapa se define que frases harán parte del corpus, la cantidad de fonemas que deben aparecer en el total del corpus, el número de frases que debe grabar cada locutor, el número de locutores necesarios, el género y edad de los locutores y las especificaciones que debe tener el lugar donde se realizarán las grabaciones.

Para obtener mejores resultados en el entrenamiento, las frases no fueron etiquetadas por medio de un segmentador automático sino que cada una de las frases ha sido examinada y segmentada manualmente por dos personas donde una es lingüista y la otra traductora en simultánea.

Se realizó la grabación de las frases en dos ambientes distintos y con características diferentes, por esto se consideran dos corpus, el primero fue etiquetado por la traductora en simultánea, se le llamará X, y a este corpus A. El segundo corpus lo realizó una persona con estudios en Fonoaudiología y con magíster en Lingüística, se le llamará Y, al corpus B.

2.1.2 Definición del corpus

Para el diseño de las frases de ambos corpus se tuvo en cuenta lo siguiente:

•Escoger frases relacionadas con un periódico en línea, el cual pretende entrenar un sistema que reconozca frases cuando se esté navegando por Internet y además para ejecutar algunos programas.

•El número de fonemas en todas las frases esté balanceado, es decir, que al entrenar se cuente con una buena cantidad de apariciones de cada fonema y no haya sobre-entrenamiento de alguno de ellos.

•Tener las suficientes muestras de cada una de las frases.

2.1.3 Colección de datos

El corpus A contiene 62 frases de un hombre y 63 frases de una mujer, las frases utilizadas son repetidas por cada uno de ellos. Estas personas pertenecen al rango entre 31 a 45 años. El corpus se compone de 26 fonemas, 105 palabras y de 63 frases no repetidas. Para este corpus se leyeron las frases sin previo entrenamiento. Se grabaron 125 emisiones para luego escoger las mejores de cada sujeto.

El corpus B consiste de nueve frases de 18 personas (nueve hombres y nueve mujeres) para un total de 216 emisiones. Las personas tienen diferentes rangos de edades esto para hacer un corpus homogéneo que no identifique un tipo de edad en especial. Los rangos van de 15 a 30 años, de 31 a 45 años y de 46 a 60 años. Las personas son nacidas en Santiago de Cali, no tienen alguna discapacidad verbal y además ninguno de ellos usa prótesis dental o aparatos en la boca.

El procedimiento para la grabación consistía en entregar a las personas un documento con las frases que debían leer diez minutos o menos antes de la grabación con el objetivo de que no repasaran la manera más adecuada de pronunciar las frases, pues la idea era que las

pronunciaran lo más natural posible, como si se tratara de una conversación común. En el momento de la grabación, las personas leen nueve frases de tres formas distintas, rápido, despacio y normal. Una vez rápido, una vez despacio y tres veces normal. De esta manera se tienen cinco repeticiones de la misma frase de las cuales se escogieron una o dos dependiendo del grupo al que pertenecen. Los grupos se dividen en entrenamiento, prueba y evaluación. Para entrenamiento se toman dos emisiones y para prueba y evaluación una emisión diferente por cada persona en cada grupo. Las emisiones que pertenecen a prueba y evaluación no han sido utilizadas para entrenar los HMMs, son totalmente nuevas para el sistema.

Las sesiones de grabación tomaron tres días en jornadas de seis horas aproximadamente. Se realizaron en un estudio de grabación de 4.20mts de largo, 2.80mts de alto y 2.80mts de ancho. El aislamiento del ruido dentro del estudio es del 100%. El equipo utilizado consiste en un computador con procesador Pentium 4, 1GB de memoria RAM, tarjeta de sonido Sound Blaster Creative© y sistema operativo Windows XP©. El software de grabación es Adobe Audition 1.5 ©. El micrófono es Behringer B-2 PRO©, Condensador, Patrón Polar Variable (Unidireccional-Omnidireccional), atenuador de 10dB y corte de bajos.

2.1.4 Segmentación fonética

Las emisiones del corpus A resultantes se enviaron a la persona X, ella se encargó de seleccionar 63 emisiones de una mujer y 62 de un hombre.

Las emisiones del corpus B resultantes de la grabación se enviaron a la persona Y. Esta persona se encargó de seleccionar una o dos de las cinco repeticiones, dependiendo del grupo al que pertenecerá la emisión.

La segmentación de ambos corpus fue realizada manualmente con Wavesurfer 1.8.3© 2005, el cual es software libre.

Finalmente, se obtuvo un corpus A de 125 emisiones y un corpus B de 216 emisiones cada uno con trascripción textual y fonética, donde cada emisión es de dos segundos aproximadamente.

2.2 Diseño del Experimento

El objetivo es evaluar la diferencia que existe cuando se entrenan HMMs con diferentes corpus que han sido segmentados por más de una persona, para esto se utilizaron dos corpus, como se mencionó anteriormente, las características de los corpus son muy diferentes, lo único que comparten es la gramática. El corpus A no se realizó en un ambiente diseñado para grabación, fue hecho en diferentes horarios y los locutores sólo fueron un hombre y una mujer. Mientras que el corpus B se realizó en un ambiente diseñado para grabación y los locutores fueron nueve hombres y nueve mujeres.

En este experimento, los eventos acústicos que se modelarán con HMM son los fonemas de las frases que se utilizaron. Los fonemas y las frases utilizadas serán descritos en la explicación del experimento.

Se debe escoger la topología para el HMM para definir el prototipo que tendrán los modelos. Conformado por:

-Número de estados

-Forma de las funciones de observación

-Probabilidad de transición entre estados

El prototipo de los fonemas que se usó en el experimento tiene las siguientes características:

El tamaño de los vectores de entrada, esta dado por: 13 coeficientes estáticos (MFCC), 13 coeficientes de desplazamiento y 13 coeficientes de aceleración.

El número de estados utilizados son 5, tres estados "efectivos", un estado de entrada y uno de salida. Se inicializan las medias en 0,0 y la varianza en 1,0.

La matriz de transiciones es de 5x5, que indica la probabilidad de pasar de un estado a otro.

2.2.1 Explicación del Experimento

El propósito del experimento consiste en demostrar la diferencia que existe cuando dos personas segmentan emisiones de voz individualmente, el entrenamiento segmentado con personas diferentes genera fronteras muy alejadas una de la otra.

Para esto, se realizaron cinco tipos diferentes de entrenamiento de los HMM, en los que lo único que se modificaba eran las emisiones utilizadas para entrenar, las cuales fueron segmentadas por distintas personas, como se muestra en la Tabla 1. En la tabla 2 se detallan las edades de los locutores y el número de emisiones utilizadas por rango de edad y género.

Tabla 1. Detalles de las emisiones utilizadas para el entrenamiento de HMMs

Tipo	Número Mujeres	Número Hombres	Etiqueta X	Etiqueta Y
I	2	3	X	X
II	3	2	X	X
III	2	2	X	X
IV	1	1	X	
V	3	3		X

Tabla 2. Participación por género y edad en los experimentos

Experimento		Tipo I	Tipo II	Tipo III	Tipo IV	Tipo V
Voz	Rango de Edad	Número de Emisiones en el Experimento				
Femenina	15-30	18	18	18	55	18
Femenina	31-45	27	27	36	0	18
Femenina	46-60	0	18	0	0	18
Masculina	15-30	0	0	35	52	18
Masculina	31-45	44*	44*	18	0	18
Masculina	46-60	18	0	0	0	18
	Total	107	107	107	107	108

En la tabla 2, el * se refiere a dos personas en el mismo rango de edad.

En todos los tipos de entrenamiento, las emisiones fueron distribuidas así:
-El 60% fue designado para entrenamiento.
-El 40% fue designado para prueba.

Las emisiones de prueba que se utilizaron para todos los experimentos fueran las mismas, en los que se encontraban emisiones de cuatro hombres y cuatro mujeres donde cada uno de ellos aportó nueve emisiones.

Dado que se utilizan diferentes corpus de entrenamiento, el número de ocurrencias por fonema no es el mismo en todos los tipos, en la Tabla 3 se detallan las ocurrencias por fonema:

Tabla 3. Número de ocurrencias de fonema por tipo

Fonema	T. I	T. II	T. III	T. IV	T. V	Fonema	T. I	T. II	T. III	T. IV	T. V
a	366	366	350	280	459	l	184	184	183	153	189
á	16	16	13	8	20	m	44	44	41	31	60
b	66	66	62	54	84	n	197	197	205	201	204
ch	17	17	21	22	12	o	240	240	230	190	268
d	96	99	98	82	111	ó	49	49	58	64	20
e	365	367	383	363	356	p	59	59	52	42	72
é	3	3	6	7	0	r	224	227	201	157	287
f	33	33	23	4	60	rr	15	15	12	6	24
g	22	22	19	16	36	s	252	251	285	289	190
i	240	240	240	217	266	t	104	106	104	84	114
í	18	18	19	16	10	u	53	53	59	57	38
j	37	38	37	28	36	ú	7	7	5	2	10
k	139	146	146	135	123	y	19	19	14	2	36

3 Resultados

Para mostrar los resultados, se hizo uso de Error Medio Absoluto mediante un programa desarrollado en Python por las autoras, éste genera el error medio por cada tipo de entrenamiento, y además por cada tipo de entrenamiento y fonema. Para generar el error por tipo, se suma la diferencia de fronteras entre la segmentación manual y la automática dividido entre el número de fonemas incluidos en la emisión. En la ecuación 1 se muestra el error medio por tipo para la frontera derecha. Para la frontera izquierda sería lo mismo, solo que tomando la frontera izquierda de las emisiones.

$$ErrorMedioDerxTipo = \sum_{i=0}^{a} \frac{1}{n} \sum_{j=0}^{n} | \, fdA - fdM \, |, \tag{1}$$

donde:
a = Número de emisiones
n = Número de fonemas incluidos en la emisión
fdA = Frontera Derecha de la emisión generada automáticamente
fdM = Frontera Derecha de la emisión generada manualmente

3.1 Gráficos

Con los resultados de cada tipo de entrenamiento, se calculó el error medio absoluto por cada uno. Como puede verse en la Fig. 1, el error más alto es el TipoIV con 838,36 ms en la frontera derecha y el menor error es el TipoI con 765,81 ms en la frontera derecha.

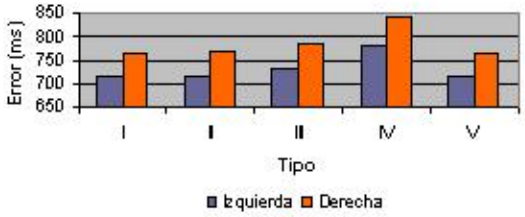

Figura 1. Error medio por Tipo de Entrenamiento.

De acuerdo a la Fig. 1, pareciera que el Tipo I y el Tipo V tienen los mismos valores, es de resaltar que en la frontera derecha el Tipo I tiene 713,80 ms y el Tipo V tiene 715,33 ms. En la frontera izquierda el Tipo I tiene 765,81 ms y el Tipo V tiene 766,08 ms.

Adicionalmente, se calculó el error medio absoluto diferenciado por tipo de entrenamiento y fonema, como se muestra en la ecuación 2.

$$ErrorMedioDerxTipoxFonema = \sum_{i=0}^{a} \frac{1}{f} \sum_{j=0}^{f} | fdA - fdM |, \quad (2)$$

donde:
a = Número de emisiones
f = Número de ocurrencias del fonema.
fdA = Frontera Derecha de la emisión generada automáticamente
fdM = Frontera Derecha de la emisión generada manualmente

Nuevamente, para la frontera izquierda sería lo mismo, solo que tomando la frontera izquierda de las emisiones.

De lo cual se obtienen dos gráficas, una por la frontera derecha de los fonemas y la segunda por la frontera izquierda, Fig. 2 y Fig. 3 respectivamente.

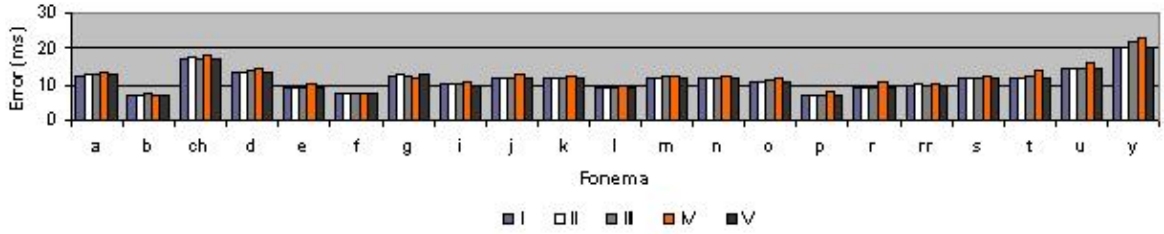

Figura 2. Error medio por Tipo de Entrenamiento y Fonema (Frontera Derecha).

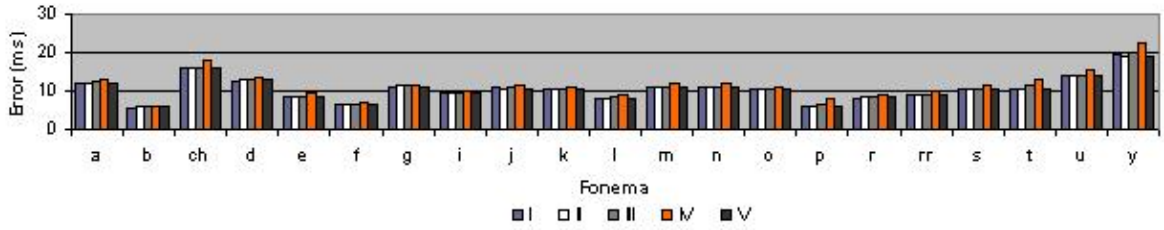

Figura 2. Error medio por Tipo de Entrenamiento y Fonema (Frontera Izquierda).

Con el objetivo de saber cuál es el tipo con mayor y menor número de errores medio por fonema, se evalúa entre los cinco tipos cual tiene el menor error medio y cual tiene el mayor error medio por fonema, por menor error se otorga un punto y por mayor error se otorga -1 punto, de esta manera se genera la Tabla 4.

Tabla 4. Número de menores y mayores errores por tipo

Tipo	I	II	III	IV	V
Puntuación	15	0	1	-20	4

Donde el Tipo I es el que menor error medio tiene, y el Tipo IV es por el contrario el que tiene la mayor cantidad de errores por fonema. Es de resaltar, que el Tipo IV fue segmentado por una sola persona y además el entrenamiento cuenta con una representación masculina y una femenina, mientras que el Tipo I fue segmentado por dos personas y tiene tres representaciones masculinas y dos femeninas. Además, el número de ocurrencias de los fonemas en dichos tipos de entrenamiento no está balanceado, lo cual no hace equitativa la evaluación, por tal razón para analizar cual tiene el mayor error medio se elimina el Tipo IV y se realiza nuevamente la puntuación, como se muestra en la Tabla 5.

Tabla 5. Número de menores y mayores errores por tipo I, II, III y V

Tipo	I	II	III	V
Puntuación	14	2	-14	2

De esta manera, el que tiene el mayor error medio es el Tipo III.

4 Conclusiones

Las conclusiones que se muestran a continuación están sujetas a las condiciones y características definidas anteriormente en este artículo.

•La segmentación manual es una característica importante ya que puede reducir o aumentar el error medio de la segmentación automática generada por el entrenamiento del corpus manual.

•El número de personas que segmentan el corpus determina sustancialmente los resultados, pues observando el error medio de los experimentos Tipo IV y V que fueron segmentados por una persona contra los Tipo I, II y III que fueron segmentados con dos, se ve una gran diferencia en los resultados obtenidos, esto se da porque los HMMs toman

características de ambos tipos de segmentación para cada fonema, y genera una discrepancia de criterios como se mencionó al principio del artículo.

•Dado que la grabación del corpus B fue desarrollado en condiciones ideales, se esperaría que fuese el que arrojara el menor error medio, sin embargo la combinación de los dos corpus A y B fue la que arrojó el menor error medio. Esto se debe principalmente a que el corpus A le aporta nuevas características al modelo de entrenamiento lo cual lo hace más robusto y completo.

•Aunque los Tipos I, II y III tienen muchas similitudes como el número de ocurrencias por fonema y que han sido segmentados por las dos mismas personas no logran los mismos resultados dado que el número de representaciones femeninas y masculinas no es el mismo, lo cual es un punto importante al momento de definir un corpus de entrenamiento.

•El Tipo IV fue eliminado del análisis pues tenía muy pocas representaciones de cada fonema y no era justo hacer un análisis con este tipo, mientras que el Tipo V que tiene suficientes ocurrencias no logra tener el menor error medio porque principalmente fue segmentado por solo una persona, a pesar de ello la diferencia con el Tipo I en la frontera izquierda es de 1,52 ms y en la frontera derecha de 0,26 ms, lo cual indica que tanto la segmentación manual por más de una persona y la cantidad de ocurrencias de cada fonema en el corpus de entrenamiento mejoran notablemente los resultados de la segmentación automática.

Agradecimientos

Las autoras reconocen las contribuciones de A. Alvarado, G. I. Álvarez, A. Bonafonte, R. San-Segundo y a J. A. Gómez por sus valiosos comentarios los cuales enriquecieron este documento. Agradecen a A. Almira y M. Camargo por su orientación lingüística y fonológica. A la Pontificia Universidad Javeriana-Cali por prestarnos los estudios de grabación de la emisora Javeriana Estéreo. A R. Jordán por el conocimiento que nos impartió, las cuales fueron las bases de esta investigación. Y a D. Linares, nuestro director de trabajo de grado por su excelente gestión y acompañamiento.

Referencias

[1]Quilis, Antonio; Fernandez, Joseph (1982). "Curso De Fonética y Fonología Españolas para Estudiantes Angloamericanos". 10ª edición, páginas 3, 4 y 9.

[2]Toledano, Doroteo; Hernandez, Luis (2003). "HMMs for Automatic Phonetic Segmentation". Speech and Audio Processing, IEEE Transactions, Volumen 11, Número 6, páginas 617 – 625.

[3]Carrillo, Aguilar (2007). "Diseño Y Manipulación de Modelos Ocultos de Markov, Utilizando Herramientas HTK. Una Tutoría". Ingeniare: Revista Chilena de Ingeniería, Volumen. 15, Número 1, páginas 18-26.

[4]López-Cózar, R; Segura, J.C.; De la Torre, A; Rubio A.J. (2001). "Una Nueva Técnica para Evaluar Sistemas Conversacionales Basada en la Generación Automática de Diálogos". Procesamiento del lenguaje natural, XVII Congreso de la SEPLN: Sociedad Española para el Procesamiento del Lenguaje Natural. Número 27, páginas 255-262.

[5]Vàzquez, Glòria; Alonso, Laura; Capilla, Joan; Castellón, Irene; Fernández, Ana (2006). "SenSem: sentidos verbales, semántica oracional y anotación de corpus". Procesamiento del Lenguaje Natural, Número 37, páginas 113-120.