

**RECONOCEDOR DE VOZ
ADAPTADO**

por

María Antonieta García Galván

Asesor de Tesis:

Dr. René O. Aréchiga Martínez

**Maestría en Ciencias de la
Computación**

**Universidad Autónoma
Metropolitana**

2005

Fecha 31 de Marzo del 2005

RECONOCEDOR DE VOZ ADAPTADO

RESUMEN

El presente trabajo consiste en el desarrollo de un sistema de reconocimiento de voz en base a Modelos Ocultos de Markov (Hidden Markov Models, HMMs) que identifica fonemas y palabras pronunciadas por un locutor hispano hablando en inglés (locutor no nativo). El sistema requiere de dos bases de datos, la de locutores nativos del inglés y la no nativa (locutores hispanos hablando inglés). La contribución de este trabajo consiste en la incorporación del conocimiento de los contrastes del Inglés y del Español para la estructura del modelo de pronunciación múltiple, permitiendo contar con modelos más robustos y eficientes.

Los sistemas de reconocimiento de voz en Inglés son entrenados con hablantes nativos del idioma, y la mayoría de los errores se atribuyen a la variedad de acentos extranjeros. Por esta razón es necesario adaptar los modelos para que adquieran los diferentes rasgos característicos de una población, en este caso de Hispanos.

Las técnicas de Adaptación se aplican a un conjunto de Modelos Ocultos de Markov bien entrenados de hablantes Independientes. Las técnicas de Adaptación resultan mejor respecto a entrenar modelos dependientes de los hablantes y usando solo una pequeña cantidad de datos de un nuevo hablante o un grupo de hablantes, se puede adaptar un conjunto de sistemas de modelos independientes del hablante para compaginar las características del nuevo hablante o grupo de hablantes.

En el presente trabajo la adaptación se realizó usando MLLR (Maximum Likelihood Linear Regression).

ABSTRACT

*This work reports the development of a speech recognition system based on Hidden Markov Models (HMMs), capable of recognizing phonemes and words spoken in English by Hispanics (non-native Speakers). The system uses two different databases, one of natives speaking English and one of non-natives (Hispanics speaking English, **Latin-American speech data base**). The contribution of this work is the incorporation of the knowledge of the contrasts between English and Spanish in the structure of the multiple pronunciation model, with the advantage of having more robust and efficient models.*

English speech recognition systems are trained with native speakers, and most of the recognition errors are attributed to the interference of foreign accents. This is why it is necessary to adapt the models to take into account the characteristic features of a given population; in this case Hispanics.

Adaptation techniques are applied to a set of well trained speaker independent Hidden Markov Models. Rather than training speaker dependent models, adaptation techniques can be applied. By using only a small amount of data from a new speaker or group of speakers, a good speaker independent system model set can be adapted to better fit the characteristics of this new speaker or group of speakers.

The adaptation technique employed was Maximum Likelihood Linear Regression (MLLR)

TABLA DE CONTENIDO

Lista de Figuras	ii
Lista de Tablas	iii
Glosario	v
Introducción	1
Capítulo I: Reconocedor de voz	4
Presentación del problema	4
Técnicas de Reconocimiento de Voz	5
Modelos Ocultos de Markov	6
Descripción del Reconocedor de voz	9
Reconocedor con palabras aisladas	11
Herramienta HTK	20
Desarrollo del Reconocedor de dígitos	25
Capítulo II: Técnicas de Adaptación	43
Método de Mapeo Espectral	44
Técnica de Mapeo de Modelado	45
Técnica de Adaptación MLLR	46
Capítulo III: Reconocedor de Voz Adaptado	48
Adaptación de los Modelos Ocultos de Markov	48
Desarrollo del Reconocedor de Voz Adaptado	49
Resultados de los Experimentos Realizados	65
Capítulo IV: Conclusiones y Trabajo Futuro	73
Conclusiones	73
Trabajo futuro	74
Bibliografía	75
Apéndice A: Resultados del Reconocedor de Voz Adaptado	76
Apéndice B: Matrices de Confusión de los Reconocedor de Frases	96
Apéndice C: Programa “promts2mlf” para transcripción a nivel fonio	114

LISTA DE FIGURAS

<i>Número</i>		<i>Página</i>
1.	Instructor Automatizado de Pronunciación de Inglés	2
2.	Parametrización de la Señal	10
3.	Entrenamiento y Reconocimiento utilizando HMMs	11
4.	Modelo de Markov “M”	13
5.	Secuencia de operaciones para calcular la variable $\alpha_t(j)$ (hacia adelante- “Forward”)	17
6.	Cálculo de la probabilidad hacia delante en función de las observaciones t y estados i .	18
7.	Secuencia de operaciones para calcular la variable $\beta_t(j)$ (hacia atrás- “Backward”)	19
8.	Etapas del procesamiento para el diseño de Sistemas de Reconocimiento de Voz con HTK	24
9.	Gramática del Reconocedor de Dígitos	27
10.	Etiquetado de la señal de voz “dígitos” con xwaves	30
11.	Proceso de adaptación de un locutor	43
12.	Método de Mapeo Espectral	44
13.	Adaptación de un hablante por la técnica de Mapeo de Modelado	45
14.	Adaptación de HMMs	47
15.	Etiquetado de la señal de voz “frases” con xwaves	55
16.	Ejemplo a del etiquetado de la señal de voz “frases” con espectro	56
17.	Ejemplo b del etiquetado de la señal de voz “frases” con espectro	57

LISTA DE TABLAS

<i>Número</i>		<i>Página</i>
1.	Reconocedor entrenado con nativos y probado con nativos	40
2.	Reconocedor entrenado con nativos y probado con no nativos	41
3.	Reconocedor entrenado con no nativos y probado con no nativos	42
4.	Resultados de los Reconocedores de dígitos	66
5.	Reconocedor entrenado con nativos del Inglés y probado con un hablante no nativo	67
6.	Reconocedor entrenado con personas nativas del idioma Inglés, adaptado con muestras de un Hispano y probado con el mismo Hispano	67
7.	Resumen de resultados de los reconocedores de frases	72

AGRADECIMIENTOS

Agradezco infinitamente a mi esposo, mis hijos y mis padres, quienes siempre me alentaron a continuar con mis estudios.

Este trabajo es de gran importancia para mí, ya que me permitió incursionar en un tema que no conocía, por lo que agradezco infinitamente la paciencia de mi asesor el Dr. René Aréchiga Martínez, con quien conté en todo momento para aclarar dudas y para orientarme en el desarrollo del proyecto, así como para transferir su conocimiento y hacer posible la realización de esta tesis.

GLOSARIO

- EM** Algoritmo de maximización de la esperanza (Expectation-Maximization Algorithm).
- Forward-backward procedure.**- Procedimiento hacia adelante – hacia atrás.
- HMM** Modelos Ocultos de Markov (Hidden Markov Models).
- HTK** Herramienta de procesamiento de voz para construir HMMs. (Hidden Markov Models Toolkit).
- LPC** Modelos de codificación de Predicción Lineal. (Linear Predictive Coding Model)
- MAP** Probabilidad Máxima a Posteriori (Maximum a posteriori).
- MLF** Archivo de etiquetas Maestro (Master Label File)
- MLLR** Regresión Lineal de máxima probabilidad. (Maximum Likelihood Linear Regression)
- MMF** Archivo Macro Maestro (Master Macro File)
- SLF** Formato de Entramado Estándar (Standar lattice Format)
- TIMIT** Base de datos que permite obtener la pronunciación en fonos de palabras en Inglés Americano.
- TMF** Archivo de modelo de transformadas (Transform Model File).

INTRODUCCION

Se han analizado diferentes artículos sobre reconocimiento de voz [1], [3],[10], entre los que destacan los desarrollados por el Instituto de Investigaciones de Stanford (SRI International) sobre la Tecnología de Reconocimiento de Voz para la enseñanza de idiomas. En dichos artículos se proporciona información de un sistema de voz de educación interactiva para apoyar el aprendizaje de idiomas. El sistema consta de 3 módulos de procesamiento que operan en serie: un extractor de rasgos de la señal de voz, el cual transforma la señal acústica en una secuencia de frecuencia espectrales; un módulo de alineación y búsqueda, el cual toma como entrada al espectro y selecciona la hipótesis de la palabra o cadena de palabras que representa mejor al espectro, y un módulo evaluador de pronunciación, el cual compara el habla no nativa con la nativa y genera una calificación.

Internacionalmente ya existen aplicaciones de reconocimiento de voz en sistemas de aprendizaje de idiomas; sin embargo, la aportación de este trabajo se basa en la orientación del sistema, ya que no tenemos conocimiento de que actualmente se haya realizado un reconocedor de voz con la aplicación de tutor del idioma Inglés para hispanos. Por otro lado, este proyecto nos permitirá crear nuestra propia tecnología de reconocimiento de voz.

Este trabajo forma parte del proyecto global **“Instructor Automatizado de Pronunciación del Inglés”** cuyo objetivo es apoyar al estudiante hispano con una herramienta interactiva que le ayude a mejorar su pronunciación del inglés [3] [4]. Las partes principales del proyecto se muestran en la Figura 1.

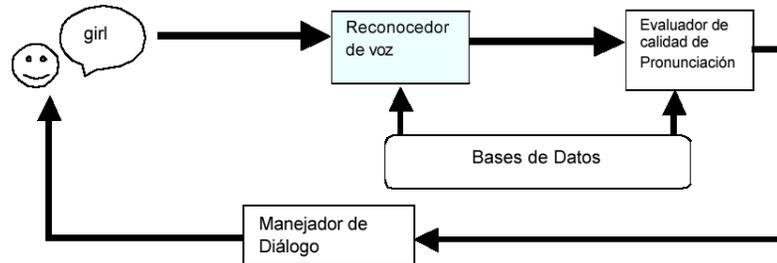


Fig. 1 Instructor Automatizado de Pronunciación del Inglés.

El Instructor Automatizado funciona de la siguiente manera: el estudiante hispano pronuncia una palabra, que será procesada por el **reconocedor de voz** con el fin de identificarla haciendo uso de modelos entrenados con las bases de datos. El resultado del reconocedor será el texto correspondiente a la mejor hipótesis de la palabra pronunciada. El evaluador de calidad de pronunciación generará una calificación que estime el grado de aproximación a la pronunciación nativa [1]. Finalmente el manejador de diálogo establecerá la relación con el estudiante para ayudarlo a mejorar su pronunciación. El Reconocedor de voz, proyecto central en este documento se subdivide en dos bloques:

Detector de pronunciación no nativa: para este módulo se cuenta con modelos de sonidos clave de vocales y consonantes que permitirán determinar si se trata de palabras nativas o no nativas. Es importante recalcar que los ciudadanos en los Estados Unidos hablan con una gran variedad de acentos dependiendo de la región geográfica en la que se encuentran o de su lugar de origen. Los sistemas de reconocimiento de voz actuales no son capaces de reconocer el acento de extranjeros [2].

Reconocedor de voz: módulo de reconocimiento de voz en base a modelos ocultos de Markov (HMMs) [5].

Bases de Datos: se requieren dos bases de datos, la de locutores nativos del inglés, que consta de palabras aisladas y frases cortas; y la no nativa (locutores hispanos hablando inglés).

Los Modelos Ocultos de Markov son una técnica de modelado probabilístico que incluyen dos componentes: una cadena de Markov de estados finitos y un conjunto finito de distribuciones de probabilidad de salida.

Capítulo I: Reconocedor de Voz

1.1 Presentación del Problema

Los sistemas de reconocimiento de voz generalmente consideran que la señal de voz es una realización de algún mensaje codificado como una secuencia de uno o más símbolos. Para efectos de operación inversa de reconocimiento de una secuencia de símbolos provenientes de la pronunciación de un hablante, la forma de onda continua de la voz primero se convierte a una secuencia de vectores parametrizados en forma discreta igualmente espaciados en el tiempo. Para que una señal pueda ser procesada digitalmente debe ser en tiempo discreto y tomar valores discretos. Si la señal a procesar es analógica, se convierte a digital haciendo un muestreo en el tiempo y obteniendo por tanto una señal en tiempo discreto y posteriormente cuantificando sus valores en un conjunto discreto.

Esta secuencia de vectores parametrizados se asume que formará una representación exacta de la forma de onda de voz basada en la duración cubierta por un solo vector (típicamente de 10 ms). La forma de onda puede ser considerada como estacionaria. A pesar de que esto no es estrictamente verdadero, es una aproximación razonable. Las representaciones paramétricas típicas comúnmente usadas son espectros alisados o coeficientes de predicción lineal además de otras representaciones derivadas de éstas.

El objetivo de los reconocedores consiste en efectuar un mapeo entre secuencias de vectores de voz y las secuencias de símbolos subyacentes requeridos.

Las dificultades que se presentan son: Primero, el mapeo de símbolos de la voz, no es uno a uno debido a que diferentes símbolos subyacentes pueden tener sonidos de voz similares. Además existen variaciones largas en las ondas de voz debidas a la variabilidad del hablante por su estado de ánimo, medio ambiente, timbre de voz, etc. Segundo, los límites entre símbolos no pueden ser identificados explícitamente de la forma de onda de la señal de voz. Por lo que no es posible tratar la forma de onda de voz como una secuencia concatenada de patrones estáticos.

El segundo problema consiste en no poder delimitar la frontera de las palabras. Esto puede evitarse restringiendo la tarea del reconocedor a solo palabras. Lo cual implica que la forma de onda de voz corresponda a un solo símbolo seleccionado de un vocabulario fijo. A pesar de este hecho que es un problema simple es un tanto artificial, no obstante tiene un amplio rango de aplicaciones. Además esto proporciona las bases para introducir ideas básicas del reconocimiento basado en Modelos Ocultos de Markov. Por esta razón, en esta tesis primero se realizaron experimentos de reconocedores de voz utilizando palabras aisladas con Modelos Ocultos de Markov (HMM).

1.2 Técnicas de Reconocimiento de Voz

Existen tres métodos sobresalientes que se enfocan al reconocimiento automático de voz con máquinas [6]:

1. El Fonético- acústico. En este método la máquina intenta decodificar la señal de voz de manera secuencial con base en características acústicas observadas de la señal y las relaciones conocidas entre las características acústicas y los símbolos fonéticos. Sin embargo, se tienen algunos problemas para tener un sistema de reconocimiento de voz exitoso ya que se requiere un conocimiento extensivo de las propiedades acústicas de las unidades fonéticas.

2. El de reconocimiento de patrones. En este método los patrones de voz se usan directamente sin determinación de características explícitas. Se tienen dos pasos: entrenamiento de patrones de voz y reconocimiento de patrones de voz a través de la comparación de patrones. Este es el método más utilizado para reconocedores de voz, debido a la simplicidad de uso, facilidad de entendimiento y por su riqueza en matemáticas y teoría de comunicaciones. **Este método es el que se utilizará para el reconocimiento de voz en este proyecto.**

3. El de Inteligencia artificial. Es un método híbrido que explota ideas y conceptos del método acústico fonético y del de reconocimiento de patrones. Involucra conceptos como el de redes neuronales.

1.3 Modelos Ocultos de Markov

Se introdujeron inicialmente a finales de la década de 1960 y principios de los años 1970. Los métodos estadísticos de los Modelos Ocultos de Markov, se han vuelto más populares en los últimos años, debido a dos razones principales: Los modelos son muy ricos en estructura matemática y pueden formarse las bases teóricas para usarse en un amplio rango de aplicaciones. Segundo, los modelos al aplicarse apropiadamente para diversas aplicaciones, trabajan muy bien en la práctica.

Los procesos del mundo real presentan salidas observables las cuales se pueden caracterizar como señales. Las señales pueden ser de naturaleza discreta (Caracteres de un alfabeto finito, vectores cuantificados de un libro de códigos, etc.); o de naturaleza continua (Muestras de voz, mediciones de temperatura, música, etc.) La fuente de la señal puede ser estacionaria (no varía con el tiempo) o no estacionarias (Las propiedades de la señal varían con el tiempo). Las señales pueden ser puras (provienen estrictamente de una sola fuente), o pueden ser afectadas por otras fuentes de señales (Ruido o por distorsiones de la transmisión, reverberaciones, etc.).

Un problema de gran interés es la caracterización de las señales del mundo real en términos de modelos de la señal. Existen diferentes razones de interés para aplicar modelos de señales. En primer lugar, un modelo de señal puede proporcionar las bases para una descripción teórica de un sistema de procesamiento de señales, los cuales pueden ser usados para procesar la señal de tal forma que proporcione una salida deseada. Por ejemplo, si se está interesado en mejorar una señal de voz que ha sido afectada por ruido y distorsión de transmisión, es conveniente obtener el modelo de la señal para diseñar un sistema el cual remueva el ruido y baje la distorsión de transmisión. Otra razón por la que los modelos de la señal son importantes, es porque ellos son potencialmente capaces de permitirnos aprender mucho acerca de la fuente de la señal (por ejemplo los procesos del mundo real, los cuales producen la señal) sin tener que tener la fuente disponible, esta propiedad es especialmente importante cuando el costo de obtener señales de la fuente actual es alto. En este caso con un buen modelo de la señal, nosotros podemos simular la fuente y aprender tanto como sea posible a través de una simulación.

Finalmente, una de las razones más importantes sobre los modelos de la señal, es que ellos nos permiten realizar sistemas prácticos, por ejemplo, sistemas de predicción, sistemas de reconocimiento, sistemas de identificación, etc. de una manera muy eficiente.

Generalmente uno puede dividir en dos los tipos de modelos de señal, en la clase de modelos determinísticos y las clases de modelos estadísticos.

Los modelos determinísticos generalmente explotan algunas propiedades específicas conocidas de la señal, por ejemplo que la señal es una onda senoidal o una suma de exponenciales, en estos casos la especificación del modelo de la señal es generalmente sencillo, todo esto es para estimar valores de los parámetros del modelo de la señal, por ejemplo, amplitud, frecuencia de la onda senoidal, amplitudes y rango de la exponencial.

La segunda clase de modelos de la señal, es el conjunto de los modelos estadísticos, en los cuales uno trata de caracterizar solo las propiedades estadísticas de la señal. Ejemplo de tales modelos estadísticos incluye los procesos Gaussianos, procesos de Poisson entre otros. La suposición destacada del modelo estadístico consiste en que la señal puede ser bien caracterizada como un proceso aleatorio paramétrico, y que los parámetros del proceso estocástico pueden ser determinados de forma precisa bien definida.

Para las aplicaciones de procesamiento de voz, tanto los modelos estocásticos como los determinísticos tienen buen éxito.

Un tipo de modelos estocásticos para la señal son los Modelos Ocultos de Markov (HMM). En los modelos Ocultos de Markov, se identifican tres problemas fundamentales para su diseño: La evaluación de la probabilidad de una secuencia de observaciones dado un Modelo Oculto de Markov específico, la determinación de la mejor secuencia de los estados del modelo y el ajuste de los parámetros del modelo, para que sea la mejor cuantificación de la señal observada.

La teoría básica de los Modelos Ocultos de Markov fue publicada en una serie de papeles clásicos por Baum y su colega a finales de la década de los años de 1960 y principios de los años 1970's y fue implementada para aplicaciones de procesamiento de voz por Baker en CMU y por Jelinek y sus colegas en IBM en los años de 1970's. Sin embargo su extensión del conocimiento y aplicaciones de la teoría de los Modelos Ocultos de Markov para el procesamiento de voz, ha

ocurrido dentro de los últimos años. Esto debido a que la teoría básica de los HMM al principio se dio a conocer en publicaciones matemáticas, y los lectores generalmente no eran ingenieros que trabajaban en problemas para las aplicaciones de procesamiento de voz.

Los tres Problemas básicos de HMM's

Problema 1: Dada la secuencia de observaciones $O = O_1 O_2 \dots O_T$, y un modelo $\lambda = (A, B, \pi)$, ¿Cómo podemos calcular eficientemente $P(O/\lambda)$, es decir, la probabilidad de la secuencia de observaciones dado el modelo?.

Problema 2: Dada la secuencia de observaciones $O = O_1 O_2 \dots O_T$, y el modelo λ , ¿Cómo seleccionamos una secuencia de observaciones que corresponde a $Q = q_1 q_2 \dots q_T$, la cual sea la óptima con sentido coherente, es decir, la que mejor especifica las observaciones.

Problema 3: ¿Cómo ajustamos los parámetros del modelo $\lambda = (A, B, \pi)$, para maximizar $P(O/\lambda)$?

Problema 1, consiste en la evaluación del problema, concretamente, dado un modelo y una secuencia de observaciones, cómo calculamos la probabilidad de que la secuencia de observaciones fue producida por el modelo. Podemos ver el problema como la cuantificación de que también un modelo dado empata con la secuencia de observaciones dadas. Este último punto de vista es extremadamente usado. Por ejemplo, si nosotros consideramos el caso en el cual estamos tratando de escoger entre diferentes modelos compitiendo, la solución al problema uno nos permite seleccionar el modelo que empata mejor con las observaciones.

Problema 2, es en el cual nosotros intentamos dejar al descubierto la parte escondida del modelo, por ejemplo, encontrar la secuencia de estados correcta. Debe ser claro, que para todos los casos degenerados, no hay una secuencia correcta para encontrarla. De ahí que para situaciones prácticas, nosotros generalmente utilicemos un criterio de optimización para resolver este problema como mejor sea posible. Desafortunadamente, como nosotros vemos existen algunos criterios óptimamente razonables que pueden ser impuestos, y de ahí que la selección del criterio de usos típicos podrían ser para aprender acerca de la estructura del modelo, para encontrar la secuencia de

estados en el reconocimiento de voz continuo o para obtener el promedio de estadísticas de estados individuales, etc.

Problema 3, Es en el que intentamos optimizar los parámetros del modelo, de tal forma que describa mejor como ocurre una secuencia de observaciones dada. La secuencia de observaciones usada para ajustar los parámetros del modelo se llama secuencia de entrenamiento debido a que es usada para ajustar los parámetros del modelo, se llama secuencia de entrenamiento debido a que es usada para entrenar los Modelos ocultos de Markov. EL problema de entrenamiento es uno de los más decisivos para la mayoría de las aplicaciones de HMM's debido a que nos permite adaptar óptimamente los parámetros del modelo para datos de entrenamiento observados, por ejemplo crear los mejores modelos para formar fonemas reales.

1.4 Descripción del Reconocedor de Voz

Los pasos básicos que se siguen en el módulo de reconocimiento son:

1. Parametrización de la señal
2. Entrenamiento de Modelos
3. Exploración de alternativas de estructura de los modelos
4. Comparación de Patrones
5. Toma de decisión

La función de la parametrización de la señal consiste en representar los eventos acústicos relevantes en la señal de voz en términos de un conjunto compacto y eficiente de parámetros de la señal. Los métodos de análisis espectral son considerados el corazón del procesamiento de la señal en un Sistema de reconocimiento de voz. Dos alternativas del análisis espectral son: el modelo de análisis espectral de banco de filtros y el modelo de análisis espectral de codificación de predicción lineal (Linear Predictive Coding Model, LPC). Otros parámetros valiosos son los aspectos dinámicos de estos anteriores así como la energía de la señal.

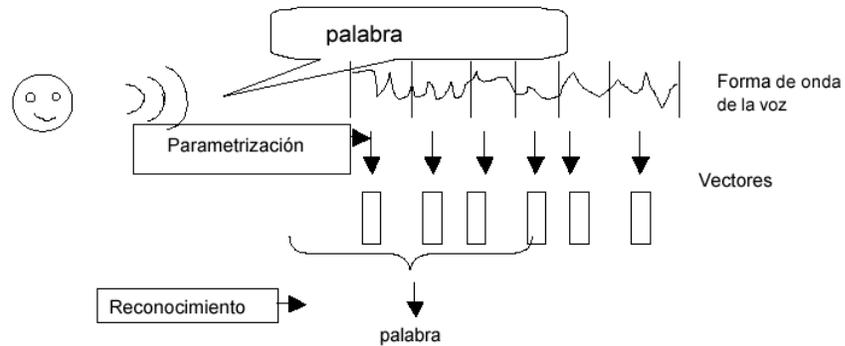


Fig. 2. Parametrización de la señal.

El entrenamiento de modelos se realizará usando los algoritmos de Baum Welch (forward-backward), el cual permite que dado un modelo λ) entrenado y una secuencia de observaciones O , calcule la probabilidad de que O fue producida por dicho modelo; este es el problema de reconocimiento $P(O|\lambda)$. La exploración de las alternativas de estructuras incluirá modelos alternos (nativos y no nativos en forma independiente), modelos adaptados (adaptación Bayesiana) y modelos compuestos que consisten de redes de fonemas que incluyen pronunciaciones nativa y no nativa. Cada palabra hablada se representa por una secuencia de observaciones: $O = o_1, o_2, \dots, o_T$

El reconocimiento de cada palabra será calcula por:

$$\arg \max_i \{P(w_i | O)\} \dots\dots\dots(1)$$

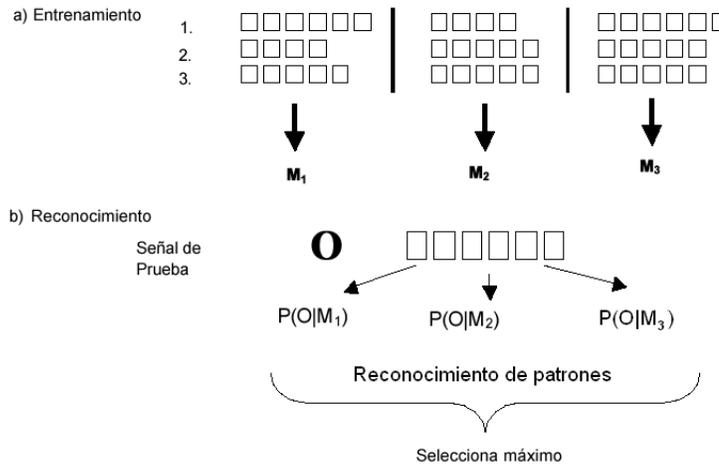


Fig. 3. Entrenamiento y Reconocimiento usando HMMs.

La comparación de patrones se requiere para evaluar el grado de similitud entre ellos. La voz es directamente representada por la secuencia de vectores espectrales, de tal forma que se define un patrón de prueba t , como la concatenación de las tramas espectrales sobre la duración de la voz, de tal forma que $t = \{t_1, t_2, t_3, \dots, t_l\}$, donde cada t_i es el vector espectral de la entrada de voz al tiempo i , y l es el número total de tramas de voz. De manera similar se define un conjunto de patrones de referencia, $\{R_1, R_2, \dots, R_V\}$ donde cada patrón de referencia, R_j , es también una secuencia de tramas espectrales tal que $R_j = \{r_{j1}, r_{j2}, \dots, r_{jJ}\}$. El propósito de la comparación de los patrones es determinar la diferencia o distancia del patrón de prueba con respecto al patrón de referencia $R_j, 1 \leq j \leq V$. La toma de decisión tiene por objeto identificar el patrón de referencia más cercano al de prueba y asociar este patrón de referencia a la señal de voz.

1.5 Reconocedor con Palabras Aisladas.

Consideremos que cada palabra aislada se representa por una secuencia de vectores de voz u observaciones O , que se definen con la siguiente expresión.

$$O = o_1, o_2, \dots, o_T \dots\dots\dots(2)$$

Donde o_T es el vector de voz observado en el tiempo t . El problema de reconocimiento de palabras aisladas puede ser considerado con el siguiente cálculo.

$$\frac{\arg \max}{i} \{P(w_i | O)\} \dots\dots\dots(3)$$

Donde w_i es la i ésima palabra del vocabulario. Esta probabilidad no es calculable directamente pero utilizando la regla de Bayes se obtiene:

$$P(w_i | O) = \frac{P(O | w_i)P(w_i)}{P(O)} \dots\dots\dots(4)$$

Esto se interpreta de la siguiente manera: para un conjunto de probabilidades dadas $P(w_i)$, la palabra pronunciada más probable depende solo de la probabilidad $P(O|w_i)$. Dada la dimensionalidad de la secuencia de observación, la estimación directa del conjunto de la probabilidad condicional $P(o_1, o_2, \dots | w_i)$ no es práctica. Sin embargo, si se asume un modelo paramétrico como el modelo de Markov, entonces la estimación de datos es posible debido a que el problema de estimación de las densidades de observación condicional de clases $P(O|w_i)$ se reemplaza por un problema más simple de estimación de los parámetros del Modelo de Markov.

El reconocimiento de voz en base a HMM, asume que la secuencia de vectores de voz observados que corresponde a cada palabra es generado por un modelo de Markov.

Un modelo de Markov es una máquina de estados finita la cual cambia de estados una vez en cada unidad de tiempo y cada tiempo t que un estado j es introducido, un vector de voz o_t se genera de la densidad de probabilidad $b_j(o_t)$. La transición de un estado i a un estado j es también probabilística y se gobierna por la probabilidad discreta a_{ij} . Como ejemplo se puede tener un modelo de cinco estados que se mueve a través de la secuencia de estados $X=1,2,3,3,4,4,5$ generando la secuencia o_1 a o_5 , como se muestra en la figura siguiente.

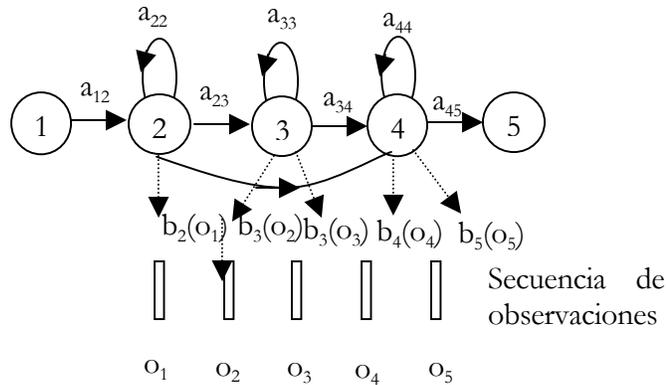


Fig. 4. Modelo de Markov "M"

La probabilidad en conjunto de que O sea generada por el modelo M moviéndose a través de la secuencia de estados X, se calcula simplemente por el producto de las probabilidades de transición y las probabilidades de salida. De tal forma que la secuencia de estados X es:

$$P(O, X | M) = a_{12}b_2(o_1)a_{23}b_3(o_2)a_{33}b_3(o_3) a_{34}b_4(o_4)a_{44}b_5(o_5) \dots\dots\dots(5)$$

En la práctica, solo la secuencia de observaciones O es conocida la secuencia de estados subyacentes X es escondida. Esta es la razón, por la que es llamada Modelo Oculto de Markov.

Debido a que X es desconocido, la probabilidad requerida es calculada a través de la sumatoria de la secuencia de todos los estados posibles $X=x(1),x(2),x(3),\dots,x(T)$, si lo escribimos en una expresión, se tiene:

$$P(O | M) = \sum a_{x(0)x(1)} \prod_{t=1}^T b_{x(t)}(O_t) a_{x(t)x(t+1)} \dots\dots\dots(6)$$

de donde x(0) es obligada a ser el estado de entrada del modelo y x(T+1) es el estado de salida del modelo. Como una alternativa a la ecuación, la probabilidad puede ser aproximada solo considerando la secuencia de estados más cercana, esto es:

$$\hat{P}(O | M) = \max_X \left\{ a_{x(0)x(1)} \prod_{t=1}^T b_{x(t)}(o_t) a_{x(t)x(t+1)} \right\} \dots\dots\dots(7)$$

A pesar de que el cálculo directo de las ecuaciones anteriores no es manejable, existen procedimientos recursibles simples, los cuales permiten que ambas cantidades se calculen eficientemente.

Dado un conjunto de modelos M_i correspondientes a palabras w_i , se resuelve la ecuación A2 y A3 y se asume que:

$$P(O | w_i) = P(O | M_i) \dots\dots\dots(8)$$

Todo esto considera que los parámetros $\{a_{ij}\}$ y $\{b_j(o_t)\}$ son conocidos para cada modelo M_i .

Se puede apreciar que en los HMM, dado un conjunto de ejemplos de entrenamiento correspondientes a un modelo particular, los parámetros de ese modelo pueden determinarse automáticamente por un procedimiento de re-estimación robusto y eficiente. Esto logra que un número suficiente de ejemplos representativos de cada palabra puedan ser coleccionados. Después un HMM se puede construir; éste implícitamente modela todas las fuentes de posibilidades inherentes en la voz real.

El proceso de reconocimiento de palabras aisladas se resume de la siguiente manera:

- 1.- Un HMM es entrenado para cada palabra del vocabulario utilizando un número de ejemplos de esa palabra. Por ejemplo el vocabulario puede ser los dígitos del 0 al 9 ("one", "two", "three", ...).
- 2.- Para reconocer alguna palabra desconocida, la probabilidad de cada modelo generado por esa palabra se calcula y el modelo más cercano identifica la palabra (la probabilidad más alta).

Introducción al Procedimiento Forward -Backward

Para calcular la probabilidad de la secuencia de observaciones, $O = O_1 O_2 \dots O_T$, dado un modelo λ , esto es $P(O|\lambda)$. La forma más sencilla de hacerlo es enumerar diferentes secuencias de estados de longitud T (el número de observaciones). Considerando una secuencia fija de estados $Q = q_1 q_2 \dots q_T$ (9)

Donde q_1 es el estado inicial. La probabilidad de la secuencia de observaciones O para la secuencia de estados anterior es

$$P(O | Q, \lambda) = \prod_{t=1}^T P(O_t | q_t, \lambda) \quad \dots\dots\dots(10)$$

Asumiendo una independencia estadística de las observaciones se obtiene por,

$$P(O | Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdot \dots \cdot b_{q_T}(O_T) \quad \dots\dots\dots(11)$$

La probabilidad de tal secuencia de estados Q se puede escribir como

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T} \quad \dots\dots\dots(12)$$

La probabilidad conjunta de O y Q , es decir, la probabilidad de que O y Q ocurran simultáneamente, es simplemente el producto de los dos términos anteriores.

$$P(O, Q|\lambda) = P(O|Q, \lambda) P(Q, \lambda) \quad \dots\dots\dots(13)$$

La probabilidad de O dado el modelo se obtiene sumando esta probabilidad conjunta sobre todas las secuencias posibles de estados dada q

$$P(O | \lambda) = \sum_{todas Q} P(O | Q, \lambda) P(Q | \lambda) \quad \dots\dots\dots(14)$$

$$= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T)$$

La interpretación del cálculo en la ecuación anterior es la siguiente: Inicialmente en el tiempo $t=1$, estamos en un estado q_1 con probabilidad π_{q_1} y genera el símbolo O_1 (en este estado) con probabilidad $b_{q_1}(O_1)$. El reloj cambia del tiempo t al $t+1$ ($t=2$) y se realiza una transición de estado q_2 desde el estado q_1 , con probabilidad $a_{q_1q_2}$ y genera el símbolo O_2 con probabilidad $b_{q_2}(O_2)$. Este proceso continúa de esta forma hasta que se realiza toda la transición de la lista, en la probabilidad $a_{q_T-1q_T}$ y genera el símbolo O_T con probabilidad $b_{q_T}(O_T)$.

El cálculo de $P(O|\lambda)$ de acuerdo a la definición de la última expresión involucra un orden de $2T \cdot N^T$ cálculos, dado que por cada $t=1,2, \dots, T$, existen N estados posibles los cuales pueden obtenerse. Por ejemplo hay N^T posibles secuencias de estado y por cada secuencia de estado se requiere aproximadamente $2T$ cálculos por cada término de la sumatoria en la última expresión.

Para ser precisos, necesitamos $(2T-1)N^T$ multiplicaciones y N^T-1 sumas. Este cálculo no resulta eficiente computacionalmente, incluso para valores pequeños de N y T . Por ejemplo, para $N=5$ estados y $T=100$ observaciones, se tiene un orden de $2 \cdot 100 \cdot 5^{100} \approx 10^{72}$ cálculos. Claramente se requiere un procedimiento más eficiente para resolver este problema. Este procedimiento se llama forward-backward (hacia delante – hacia atrás).

Procedimiento forward-backward.

Consideremos las siguientes variables $\alpha_t(j) = P(O_1 O_2 \dots O_t, q_t = S_j | \lambda)$ la probabilidad de la secuencia de observaciones, $O_1 O_2 \dots O_t$ (hasta el tiempo t), y estado S_j al tiempo t , dado el modelo λ . Se puede resolver inductivamente como sigue.

1) Inicialización:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N. \quad \dots\dots\dots(15)$$

Este paso inicializa la probabilidad hacia adelante para la probabilidad conjunta de los estados S_j y la Observación inicial O_1 .

2) Inducción:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1 \quad \dots\dots\dots(16)$$

$$1 \leq j \leq N$$

El paso de inducción, es el corazón de los cálculos hacia adelante. Ver figura 5, en la que se muestra como el estado S_j alcanza al tiempo $t+1$ de N posibles estados, S_i , $1 \leq i \leq N$, al tiempo t . Debido a que $\alpha_t(i)$ es la probabilidad del evento conjunto de que $O_1 O_2 \dots O_t$ son observadas, y el estado al tiempo t es S_i , el producto $\alpha_t(i)a_{ij}$ es entonces la probabilidad del evento conjunto de que $O_1 O_2 \dots O_t$ son observadas y el estado S_j , es alcanzado al tiempo $t+1$ a través del estado S_i en el tiempo t . Resumiendo, este producto sobre todos los N estados posibles S_i , $1 \leq i \leq N$ al tiempo t resulta la probabilidad de S_j al tiempo $t+1$ con todo el acompañamiento previo de las observaciones parciales. Una vez que se conocen éstas y se conoce S_j es fácil notar que $\alpha_{t+1}(j)$ se obtiene estimando las observaciones O_{t+1} en el estado j , por ejemplo multiplicando la cantidad sumada por la probabilidad $b_j(O_{t+1})$. El cálculo de la expresión 16 se realiza para todos los estados j , $1 \leq j \leq N$ para una t dada, los cálculos son iterados para $t=1,2, \dots, T-1$.

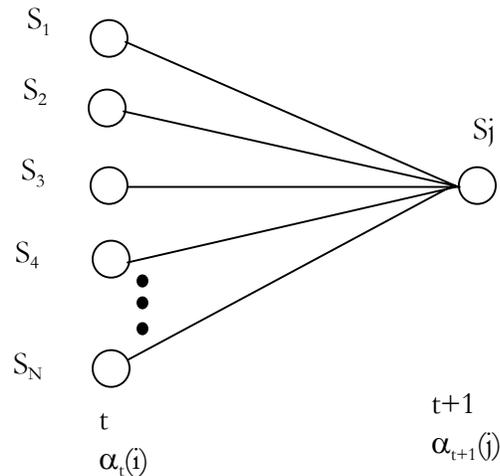


Fig. 5 Secuencia de operaciones para calcular la variable $\alpha_t(j)$ (hacia adelante-“Forward”)

3) Terminación:

Genera el cálculo deseado de $P(O|\lambda)$ como la suma en la terminal de la variable hacia delante

$$\alpha_T(i) = P(O_1 O_2 \dots O_T, q_T = S_i | \lambda) \dots\dots\dots(17)$$

Y de ahí $P(O|\lambda)$ es justo la suma de las $\alpha_T(i)$'s .

Si examinamos los cálculos involucrados en el cálculo de $\alpha_t(j)$, $1 \leq t \leq T$, $1 \leq j \leq N$, podemos observar que se requiere cálculos del orden de N^2T , en lugar $2TN^T$ como se requeriría a través del cálculo directo. Y siendo mas preciso se requiere $N(N+1)(T-1) + N$ multiplicaciones y $N(N-1)(T-1)$ sumas. Para $N=5$, $T=100$, necesitaríamos aproximadamente 3000 cálculos para el método hacia delante, respecto a 10^{72} cálculos en el método directo, ahorrando ≈ 69 órdenes de magnitud. Los cálculos de la probabilidad hacia delante se basan en el entramado que se muestra en la figura 6. La clave está en que debido a que existen solo N estados (nodos en cada ranura de tiempo en el enmallado), todas las secuencias posibles de estado se reunirán dentro de N nodos, no importando el tamaño de la secuencia de observaciones. Al tiempo $t=1$ (primer ranura del enmallado), se necesita calcular valores de $\alpha_1(i)$, $1 \leq i \leq N$. En los tiempos $t=2,3, \dots, T$, solo se necesita calcular valores de $\alpha_t(j)$ $1 \leq j \leq N$, donde cada cálculo involucra solo N valores previos de $\alpha_{t-1}(j)$ debido a que cada uno de los N puntos del cuadrículado (rejillas) es alcanzado desde el mismo punto N del cuadrículado previo al ranura de tiempo.

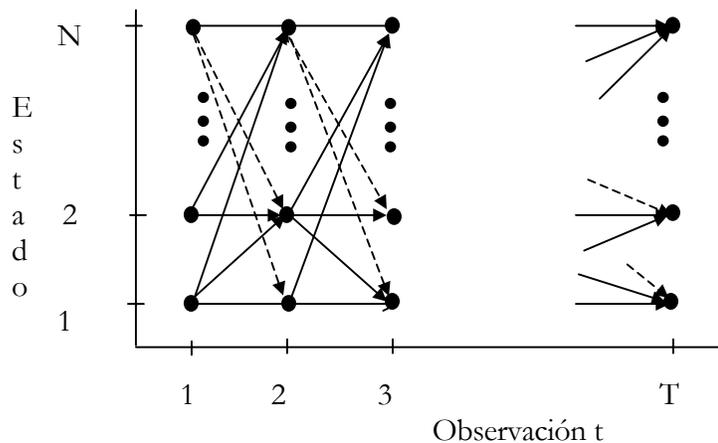


Fig. 6 Cálculo de la probabilidad hacia delante en función de las observaciones t y estados i .

De manera similar, se puede considerar una variable hacia atrás $\beta_t(i)$ definida como:

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T | q_T = S_i, \lambda) \dots\dots\dots(18)$$

Por ejemplo, para calcular la probabilidad de la secuencia de observaciones parciales de $t+1$ al final, dado el estado S_j en el tiempo t y el modelo λ . Nuevamente podemos resolver $\beta_t(i)$ inductivamente, como sigue:

1) Inicialización:

$$\beta_T(i) = 1, 1 \leq i \leq N. \dots\dots\dots(19)$$

Esta inicialización arbitrariamente define a $\beta_T(i)$ como 1 para toda i .

2) Inducción:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t=T-1, T-2, \dots, 1, 1 \leq i \leq N \dots\dots\dots(20)$$

Considera, que de haber estado en S_j en el tiempo t , y considerando a la secuencia de observaciones desde el tiempo $t+1$, se tienen que considerar todos los estados posibles S_j al tiempo $t+1$, contabilizando todas las transiciones de S_i a S_j (los términos a_{ij}), así como las observaciones O_{t+1} en el estado j (los términos $b_j(O_{t+1})$), y después estimar el resto de la secuencia de observaciones parciales del estado j (de los términos de $\beta_{t+1}(j)$). Tanto el método hacia delante como el método hacia atrás son de gran utilidad para resolver los problemas 2 y 3 de los Modelos Ocultos de Markov.

El cálculo de $\beta_t(i)$, $1 \leq t \leq T$, $1 \leq i \leq N$, requiere del orden de N^2T cálculos, y puede ser calculado en la estructura de enmallado de forma similar a la figura 5.

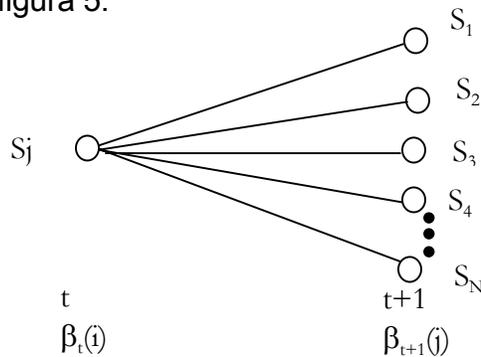


Fig. 7 Secuencia de operaciones para calcular la variable $\beta_t(i)$ (hacia atrás-“Backward”)

1.6 Herramienta HTK

Esta herramienta se diseñó con el propósito principal de construir Modelos Ocultos de Markov para el procesamiento de voz, específicamente para el diseño de reconocedores de voz.

Como primer paso las herramientas del HTK se utilizan para estimar los parámetros del conjunto de HMMs utilizando archivos de entrenamiento que contienen la pronunciación y su transcripción asociada. Posteriormente las pronunciaciones desconocidas se transcriben a través de las herramientas del HTK para su reconocimiento.

La mayoría de las funcionalidades del HTK están construidas en módulos de librerías. Las herramientas del HTK están diseñadas para correr bajo el estilo de comandos en línea. Proporcionando el nombre de la herramienta, los archivos y argumentos opcionales; los valores opcionales están siempre separados del nombre de la opción por un espacio. Como ejemplo consideremos la siguiente herramienta hipotética:

```
HFxx -T 1 -f 34.3 -a -s archivo archivo1 archivo2
```

Se tienen dos argumentos principales `archivo1` y `archivo2` y cuatro argumentos opcionales. Las opciones siempre se escriben con el nombre de la opción que corresponde a una letra seguida del valor de la opción. Por ejemplo la opción `-f` es un número real, el valor de la opción `T` es un número entero y el valor de `-s` es una cadena. En el caso de la opción `-a` no sigue ningún valor, ya que corresponde a una bandera para habilitar o deshabilitar alguna característica de la herramienta. Las opciones cuyo nombre es una letra mayúscula tienen el mismo significado para todas las herramientas. Por ejemplo `-T` se utiliza para controlar la ubicación de salida de la herramienta HTK. La operación de la herramienta se puede controlar por parámetros almacenados en un archivo de configuración, por ejemplo:

```
HFxx -C config -f 34.3 -a -s archivo archivo1 archivo2
```

En este caso a través de la opción `-C`, la herramienta hipotética HFxx cargará los parámetros que se encuentran en el archivo de configuración llamado `config` durante el procedimiento de inicialización.

Preparación de datos

Para construir el conjunto de HMMs , debemos tener un conjunto de archivos de datos (señales de voz) en el formato apropiado y su transcripción asociada. Así mismo se deben tener los archivos etiquetados.

En caso de requerir realizar el grabado de los archivos de voz, se puede utilizar la herramienta HSLab.

La herramienta HCopy se utiliza para parametrizar los datos, se utiliza para copiar uno o más archivos fuentes en un archivo de salida (concatena archivos), colocando las variables de configuración apropiadas.

HList se puede usar para checar el contenido de cualquier archivo de voz y se puede usar para verificar los resultados de cualquier conversión antes de procesar una gran cantidad de datos.

La herramienta HLEd es un editor de etiquetas diseñado para realizar las transformaciones requeridas de los archivos etiquetados. Esta herramienta puede colocar los archivos de salida en un solo Archivo de etiquetas maestro “MLF” (Master Label File), el cual es conveniente para los procesos posteriores.

Herramientas de Entrenamiento

El siguiente paso en la construcción del sistema consiste en definir la topología requerida para cada HMM, escribiendo una definición del prototipo. Se puede utilizar cualquier topología deseada, la definición del HMM se puede almacenar en forma externa como un texto simple a través de un editor de texto. El propósito de la definición del prototipo es solo especificar las características generales de la topología del HMM. Los parámetros serán calculados posteriormente por las herramientas de entrenamiento. El proceso de entrenamiento, se lleva a cabo en etapas. Primero se debe crear un conjunto de modelos inicial.

La herramienta Hlnit y HRest proporcionan un entrenamiento del estilo de palabras aisladas. Cada uno de los HMMs requeridos se genera individualmente. Hlnit lee en todos los datos de entrenamiento y deja todos los ejemplos del fonio requerido. Después calcula iterativamente un conjunto inicial de valores de parámetros utilizando el procedimiento k-media segmentada. En el primer ciclo, los datos de entrenamiento están segmentados uniformemente, cada estado del modelo se empata con el correspondiente segmento de dato y se estima la media y la varianza. En el segundo ciclo, así como en los

sucesivos, la segmentación uniforme se reemplaza por el alineamiento usando el algoritmo de Viterbi. Los valores de los parámetros iniciales calculados por HInit son mas adelante reestimados por la herramienta HRest. Nuevamente los datos etiquetados se utilizan, pero esta vez el procedimiento k-media segmentada se reemplaza por el procedimiento de reestimación Baum-Welch.

Una vez creado el conjunto inicial de modelos, la herramienta HERest se utiliza para realizar un entrenamiento integrado utilizando el conjunto de entrenamiento completo. HERest realiza una simple reestimación Baum Welch del conjunto total de modelos de fonios HMM simultáneamente. Para cada expresión de entrenamiento, los modelos de fonios correspondientes se concatenan y después el algoritmo forward-backward se utiliza para acumular las estadísticas de ocupación de estados, medias, varianzas, etc. para cada Modelo Oculto de Markov en la secuencia. Cuando se han procesado todos los datos de entrenamiento, las estadísticas acumuladas se utilizan para calcular la reestimación de los parámetros de los Modelos ocultos de Markov.

La filosofía de la construcción de un sistema en HTK consiste en que los HMMs deben ser refinados incrementalmente. Es decir una progresión típica se inicia con un simple conjunto de Gaussianas de modelos de fonios independientes del contexto, y después iterativamente se refinan expandiéndolas para incluir dependencia del contexto y usar múltiples mezclas de componentes de distribución Gaussianas. La herramienta HHEd es un editor de definición de HMM la cual clonará los modelos en conjuntos dependientes del contexto, aplicando una variedad de parámetros uniendo e incrementando el número de componentes mezclados en distribuciones específicas. El proceso consiste en modificar un conjunto de HMMs en estados utilizando HHEd y después reestimar los parámetros del conjunto modificado utilizando HERest después de cada estado. Para mejorar el desempeño para hablantes específicos, la herramienta HEAdapt y HVite se puede utilizar para adaptar HMMs a mejores modelos de características de hablantes particulares utilizando pequeñas cantidades de datos de entrenamiento o adaptación. El resultado final será un sistema adaptado.

Herramienta de Reconocimiento.

HTK proporciona una herramienta de reconocimiento llamada HVite, la cual usa un algoritmo de 'token passing' para realizar el reconocimiento de voz basado en Viterbi. HVite toma como entrada una red describiendo la secuencia de palabras permitida, la definición del diccionario definiendo cómo se debe pronunciar cada palabra y un conjunto de HMMs. Su forma de operación consiste en convertir la red de palabras para una red de fonios y después unir a la definición de HMM adecuada para cada instancia de fonio. El reconocimiento se realiza para un conjunto de archivos de voz almacenados.

Las redes de palabras necesarias para manejar HVite son generalmente ciclos (loops) simples de palabras en las cuales cualquier palabra puede seguir de cualquier otra. La herramienta HParse se utiliza para convertir el lenguaje de especificación de la gramática en una notación equivalente de red de palabras. Finalmente se requiere de la construcción de un diccionario a través de la herramienta HDMan.

Herramienta de Análisis

Cuando ya se ha construido un reconocedor con base en HMM, se necesita evaluar su desempeño. Esto generalmente se realiza utilizando el reconocedor para transcribir algunas oraciones de prueba pregrabadas y conjuntando la salida del reconocedor con las referencias correctas de las transcripciones. Esta comparación se realiza con la herramienta Hresults, la cual utiliza programación dinámica para alinear las dos transcripciones, después cuenta errores de sustituciones, eliminaciones e inserciones. Esta herramienta presenta la opción de generar una matriz de confusión.

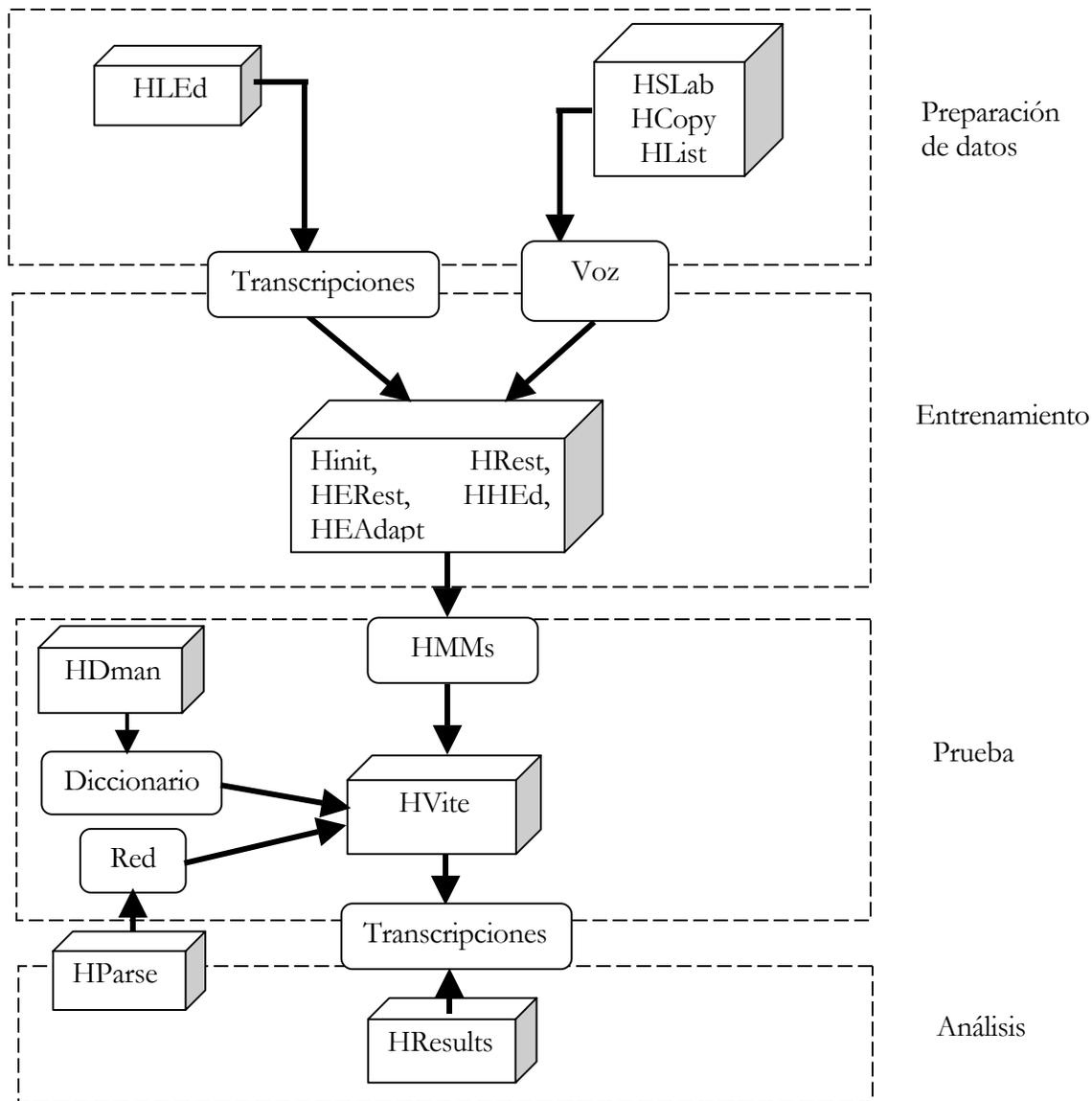


Fig. 8 Etapas de Procesamiento para el diseño de Sistemas de Reconocimiento de Voz con HTK

1.7 Desarrollo del Reconocedor de Dígitos

Representamos la señal de voz de una palabra dada como una secuencia de tiempos de vectores espectrales codificados. La codificación se realiza con M vectores espectrales únicos, de ahí que cada observación es el índice del vector espectral más cercano para la señal de voz original. De este modo, para cada palabra del vocabulario, se tiene una secuencia de entrenamiento. Esta consiste en un número de repeticiones de secuencias de índices de códigos de la palabra, estas repeticiones fueron tomadas de diferentes hablantes hispanos pronunciando en inglés los dígitos decimales. La primer tarea consiste en construir modelos de palabras individuales (los dígitos en este caso).

Como primera parte del proyecto se realizaron diferentes Reconocedores de dígitos. Los reconocedores de Voz se realizaron en una computadora PC con sistema operativo redhatLINUX 6.2. Como herramienta para el diseño de reconocedores de voz, se utilizó el software HTK V2.2.

Para la realización de estos reconocedores, se utilizaron los siguientes directorios: 'tidata': En el cual se colocaron 7 archivos de entrenamiento

Archivos de forma de onda	Archivos etiquetados	Transcripción de los archivos
tr1.htk	tr1.lab	tr1.phn
tr2.htk	tr2.lab	tr2.phn
...
tr7.htk	tr7.lab	tr7.phn

Directorio 'networks' en el que se incluye la definición de la red que para los reconocedores de dígitos el archivo se llamó 'monLattice' y la gramática de la red "monNetwork".

Directorio 'accs' el cual contiene subdirectorios en los que se guardan archivos producidos por la herramienta HERest. Subdirectorio hmm.2

Directorio 'data' el cual contiene subdirectorios en donde se colocarán las versiones parametrizadas de los datos.

Directorio 'hmms' el cual contiene subdirectorios dentro de los cuales se colocarán hmms en varios estados de entrenamiento. Subdirectorios hmm.0, hmm.1, hmm.2 (contiene los prototipos EIGHT FIVE, FOUR, NINE, ONE, SEVEN, SI, L SIX TEN, THREE, TWO, ZERO)

Directorio 'lists' el cual contiene varios archivos: vocabulario, datos de entrenamiento y prueba, etc (bcplist, dataList1, dataList2, dataList3.

Directorio 'proto' en el cual se colocarán los prototipos HMM requeridos.

Directorio 'configs' el cual contiene un archivo de configuración (monPlainM1S3.dcf).

Directorio 'results' el cual contiene los archivo de resultados para cada configuración encontrada en el directorio 'configs'

Directorio 'labels' en el cual se tienen subdirectorios dentro de los cuales se colocan los archivos etiquetados con formato HTK lbi, mon, mon-hisp, mon-orig, mon_h_h rbi, tri).

Directorio 'test' donde se colocan los archivos reconocidos a través de la herramienta HVite.

El Procedimiento para la realización de los Reconocedores de dígitos (Reconocedor de Palabras) se establece en los siguientes pasos:

Preparación de Datos

Se realizó el archivo ("monNetwork") en la que se especificó la gramática del reconocedor de dígitos. La gramática es la base con la que se desarrollará la red de reconocimiento, es la que establece las palabras que reconocerá y la secuencia en la que se debe dar.

Contenido del archivo "monNetwork"

```
$digit = SIL | ONE | TWO | THREE | FOUR | FIVE | SIX | SEVEN |  
EIGHT | NINE | TEN | ZERO;  
(<$digit>)
```

Representando con un diagrama la gramática anterior tendríamos la siguiente figura.

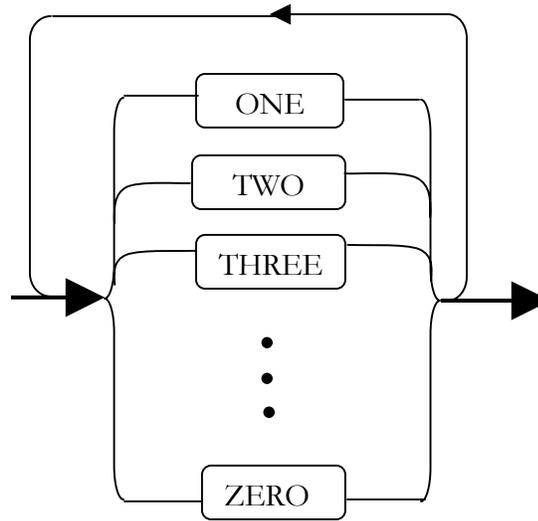


Fig. 9 Gramática del Reconocedor de Dígitos.

Una vez establecida la gramática se obtiene la red equivalente en el formato de entramado estándar (SLF) HTK en el archivo monLattice utilizando la herramienta HParse.

HParse monNetwork monLattice

Contenido del archivo “monLattice”

```

VERSION=1.0
N=15 L=37
I=0 W=ZERO
I=1 W=!NULL
I=2 W=TEN
I=3 W=NINE
I=4 W=EIGHT
I=5 W=SEVEN
I=6 W=SIX
I=7 W=FIVE
I=8 W=FOUR
I=9 W=THREE
I=10 W=TWO
I=11 W=ONE
I=12 W=SIL
I=13 W=!NULL
I=14 W=!NULL
J=0 S=0 E=1
J=1 S=1 E=13
J=2 S=1 E=12
J=3 S=1 E=11
J=4 S=1 E=10
J=5 S=1 E=9
J=6 S=1 E=8
J=7 S=1 E=7
J=8 S=1 E=6
J=9 S=1 E=5
J=10 S=1 E=4
J=11 S=1 E=3
J=12 S=1 E=2
J=13 S=1 E=0
J=14 S=2 E=1
J=15 S=3 E=1
J=16 S=4 E=1
J=17 S=5 E=1
J=18 S=6 E=1
J=19 S=7 E=1
J=20 S=8 E=1
J=21 S=9 E=1
J=22 S=10 E=1
    
```

J=23	S=11	E=1	J=30	S=14	E=7
J=24	S=12	E=1	J=31	S=14	E=6
J=25	S=14	E=12	J=32	S=14	E=5
J=26	S=14	E=11	J=33	S=14	E=4
J=27	S=14	E=10	J=34	S=14	E=3
J=28	S=14	E=9	J=35	S=14	E=2
J=29	S=14	E=8	J=36	S=14	E=0

Una red de palabras en SLF consiste en una lista de nodos y arcos. En este archivo se establece en la primera línea el número de versión; en la segunda línea se define el tamaño de la red: Número de nodos (N) y número de arcos (L). Los nodos representan palabras y los arcos las transiciones entre palabras. En los siguientes renglones del archivo se indica la lista de nodos (I, número de nodo) y la palabra (W). Y por último se listan los arcos y nodos. Cada definición de un nodo y un arco se escribe en una línea, en donde J es el número de nodo, S es el nodo de inicio y E es el nodo de salida.

Además de la red, necesitamos crear un diccionario, para lo cual requerimos de un archivo (“bcplist”) conteniendo la lista de palabras requeridas en el para el entrenamiento.

Contenido del archivo “bcplist”

SIL
EIGHT
FIVE
FOUR
NINE
ONE
SEVEN
SIX
TEN
THREE
TWO
ZERO

También se utilizó un diccionario de pronunciación de Inglés Americano (dict_timit) para generar un diccionario específico indicando, para cada palabra requerida en el reconocimiento, su pronunciación en fonios. Para crear el diccionario de esta aplicación se utilizó el siguiente comando.

```
HDMAN -m -w bcplist -n monophones0 dlog dict dict_timit
```

Archivo de salida dict

EIGHT	ey t sp
FIVE	f ay v sp
FOUR	f ao r sp
NINE	n ay n sp
ONE	w ah n sp
SEVEN	s eh v ax n sp
SEVEN	s eh v n sp
SIL	sil
SIX	s ih k s sp
TEN	t eh n sp
THREE	th r iy sp
TWO	t uw sp
ZERO	z ih r ow sp

Archivo monophones0

ey	s
t	eh
sp	ih
f	k
ay	th
v	iy
ao	uw
r	z
n	ia
w	ow
ah	

Archivos de datos: se utilizaron archivos previamente grabados por 15 personas diferentes del sexo masculino cuyo idioma nativo es el Inglés (Americano), tomados de la Base de datos TIMIT, pronunciando los dígitos en inglés. Así mismo se contó archivos grabados de 18 personas distintas del sexo masculino cuyo idioma nativo es el español, pronunciando los dígitos en inglés.

Para realizar el grabado de un locutor en cualquier momento, se puede utilizar la herramienta HSLab del HTK, que a través de su menú,

permite realizar el grabado (generando el archivo de la señal) y es también una herramienta para etiquetar la señal.

A través de esta herramienta diferentes hablantes pueden grabar la pronunciación de las frases requeridas para el entrenamiento o prueba del reconocedor. Los archivos generados son con extensión .htk

Transcripción de los archivos: Para el entrenamiento de los Modelos Ocultos de Markov, se requiere asociar cada archivo de entrenamiento con una transcripción a nivel fonio. Primeramente es necesario tener un archivo etiquetado por cada archivo de señal de voz., para etiquetar los archivos se usa un software (xwaves, HSLab o ENSIG) para editar la señal de voz de cada archivo, seleccionar los fragmentos de las señales en los que no se percibe sonido (señal) y etiquetarlos como silencios, determinar los fragmentos de la señal de voz delimitando palabra por palabra e indicando con una etiqueta el nombre de la palabra pronunciada, para realizar el etiquetado de archivos de una manera más precisa, se utilizó el espectro de la señal y se le incluyeron los formantes de dicha señal, lo cual permitió delimitar mejor la frontera de cada palabra.

Cuando se utilizan los archivos MLF, HTK busca para cada archivo con extensión .htk su archivo etiquetado con el mismo nombre pero con extensión .lab.

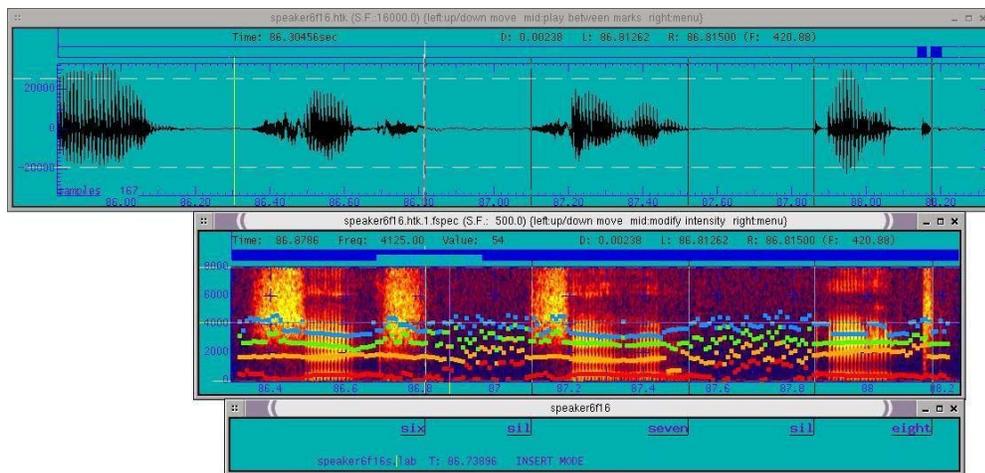


Fig. 10. Etiquetado de la señal de voz “digitos” con xwaves. Para realizar el etiquetado de los archivos de voz de los diferentes hablantes, se utilizó el software xwaves. Una vez que se editaron las

señales, se procedió a establecer las etiquetas de cada fragmento de la señal para etiquetar cada uno de los dígitos pronunciados por los hablantes, para esto se estuvo escuchando cuidadosamente cada archivo grabado a través de las bocinas de la computadora y se obtuvo a través del software xwaves el espectro de la señal mostrando los formantes, esto ayudó a que se delimitara de manera mas precisa las fronteras de cada palabra, los silencios y ruidos que presentaba la señal. El etiquetado se realiza con este software, ya que permite seleccionar fragmentos de la señal, reproducirlos y permite con su editor escribir la etiqueta que se asigna a ese fragmento de señal. En la figura anterior podemos observar un ejemplo, para uno de los archivos, este proceso de etiquetado básicamente se realiza en forma manual para cada archivo de voz de los diferentes locutores.

Para la transcripción es necesario tener un archivo con las palabras o frases pronunciadas de cada archivo y obtener la transcripción a nivel palabra. Esto es, obtener un archivo en el que se indique la trayectoria de cada archivo etiquetado y las palabras que contiene cada archivo, pero cada palabra debe ser escrita en un solo renglón y al final de la frase se debe terminar con un punto. Para realizar la transcripción a nivel palabra se utilizó el programa prompts2mlf del apéndice C.

A continuación se muestra el archivo MLF a nivel palabra para un hablante.

```
#!MLF!#  
"/home/ggma/reconoc/speakers/speaker6f16.lab"  
ONE  
TWO  
THREE  
FOUR  
FIVE  
SIX  
SEVEN  
EIGHT  
NINE  
TEN  
ZERO  
.
```

Una vez que los archivos MLF a nivel palabra se han creado, se pueden generar los archivos a nivel fonio para cada archivo de voz. Esto se realiza con la herramienta HLEd de la siguiente manera:

HLEd -l '*' -d dict -i phones0.mlf mkphones0.led words.mlf

Considerando que el archivo MLF a nivel palabra se encuentra almacenado en words.mlf, la transcripción a nivel fonio se almacenará en phones.mlf, la opción -l se utiliza para generar la trayectoria de salida y el archivo mkphones0.led contiene los siguientes comandos:

EX

IS sil sil

DE sp

El comando "EX" reemplaza cada palabra en "words.mlf" por su pronunciación correspondiente en el diccionario dict . El comando "IS" inserta el modelo de un silencio al inicio y fin de cada expresión Y el comando "DE" borra las etiquetas de pausa corta "sp"

```
#!MLF!#
"/speaker6f16.lab"
sil eh
w v
ah n
n ey
t t
uw n
th ay
r n
iy t
f eh
ao n
f z
ay ia
v r
s ow
ih sil
k .
```

La última parte de la preparación de datos corresponde a la parametrización de las ondas de voz en una secuencia características de vectores.

Se utilizó Mel Frequency Cepstral Coefficients (MFCCs), los cuales son derivados de FFT-based log spectra. La codificación se realiza con la herramienta configurada para convertir sus entradas en vectores

MFCC. Para lo cual se requiere un archivo de configuración (configco) en el cual se especifican los parámetros de conversión.

Contenido del archivo configco

```
#coding parameters
TARGETKIND = MFCC_E
TARGETRATE = 100000.0
SAVECOMPRESSED = T
SAVEWITHCRC = T
WINDOWSIZE = 250000.0
USEHAMMING = T
PREEMPCOEF = 0.97
NUMCHANS = 26
CEPLIFTER = 22
NUMCEPS = 12
ENORMALISE = V
```

En este archivo se especifica que los parámetros serán MFCC, el periodo de la trama es 10ms (HTK utiliza unidades de 100ns). La salida será en formato comprimido y se realizará un chequeo de error checksum CRC. El FFT debe usar una ventana Hamming y la señal debe tener un preenfasis de primer orden utilizando un coeficiente de 0.97. El banco de filtros debe tener 26 canales y 12 coeficientes MFCC deben ser de salida. La variable ENORMALISE tiene un valor por default de verdadera y realiza una normalización de energía en los archivos grabados de audio. Esta variable no puede utilizarse con audio en vivo, en cuyo caso debe ser puesta en Falso.

La parametrización se logra con la herramienta HCopy, sin embargo, se requiere de una lista de los archivos fuentes indicando su trayectoria, así como el archivo correspondiente de salida. Esta lista de archivos se almacena en un archivo referenciado como script y por convención se le agrega la extensión .scp. En la definición de la herramienta HCopy este archivo será especificado anteponiendo la opción -S y el archivo será leído como extensión del comando en línea. Para los reconocedores de dígitos, el archivo script utilizado se referenció como codetr.scp.

```
HCopy -T 1 -C configco -S codetr.scp
```

La opción -T habilita banderas "Trace" (rastreo) con la opción 1 se habilita el reporte de progreso básico.

Ejemplo del contenido del archivo codetr.scp

```
/home/ggma/reconoc/speakers/speaker10f16.htk /home/ggma/reconoc/speakers/speaker10f16.mfc  
/home/ggma/reconoc/speakers/speaker11f16.htk /home/ggma/reconoc/speakers/speaker11f16.mfc  
/home/ggma/reconoc/speakers/speaker6f16.htk /home/ggma/reconoc/speakers/speaker6f16.mfc  
/home/ggma/reconoc/speakers/speaker7f16.htk /home/ggma/reconoc/speakers/speaker7f16.mfc  
/home/ggma/reconoc/speakers/speaker8f16.htk /home/ggma/reconoc/speakers/speaker8f16.mfc  
/home/ggma/reconoc/speakers/speaker9f16.htk /home/ggma/reconoc/speakers/speaker9f16.mfc
```

Para construir los HMMs de las palabras, en este caso de los dígitos primeramente se utilizó la herramienta Hinit para calcular los parámetros iniciales de los modelos de cada una de las palabras (SIL, ONE,.. TEN) y después se utilizó Hrest para refinar los parámetros de los modelos, utilizando una reestimación con el algoritmo de Baum-Welch.

Llamado de la herramienta Hinit para HMM ONE

```
HInit -A -i 10 -L labels/bcplabs/mon -l ONE -o ONE -C  
toolconfs/hinit.conf -D -M hmms/hmm.0 -T 1 proto/ONE  
data/train/tr1.mfc data/train/tr10.mfc data/train/tr11.mfc  
data/train/tr12.mfc data/train/tr13.mfc data/train/tr14.mfc  
data/train/tr15.mfc data/train/tr2.mfc data/train/tr3.mfc  
data/train/tr4.mfc data/train/tr5.mfc data/train/tr6.mfc  
data/train/tr7.mfc data/train/tr8.mfc data/train/tr9.mfc
```

Llamado de la herramienta Hinit para HMM NINE

```
Calling HInit for HMM NINE  
HInit -A -i 10 -L labels/bcplabs/mon -l NINE -o NINE -C  
toolconfs/hinit.conf -D -M hmms/hmm.0 -T 1 proto/NINE  
data/train/tr1.mfc data/train/tr10.mfc data/train/tr11.mfc  
data/train/tr12.mfc data/train/tr13.mfc data/train/tr14.mfc  
data/train/tr15.mfc data/train/tr2.mfc data/train/tr3.mfc  
data/train/tr4.mfc data/train/tr5.mfc data/train/tr6.mfc  
data/train/tr7.mfc data/train/tr8.mfc data/train/tr9.mfc
```

En la herramienta Hinit la opción `-A` provoca que los argumentos de la línea de comando se impriman. La opción `-i 10` establece el número máximo de ciclos de estimación, en caso de omitir la opción se utiliza por default 20. La opción `-L labels/bcplabs/mon`, establece la trayectoria donde se encuentran los archivos etiquetados, por default busca en el directorio actual. La opción `-l NINE` establece el nombre

de la etiqueta del segmento, Hlnit busca en todos los archivos de entrenamiento y selecciona todos los segmentos con la etiqueta establecida. Cuando no se establece esta opción se considera que el archivo tiene un solo símbolo. La opción `-o NINE` indica el nombre que se le asignará al Modelo Oculto de Markov de salida, por default se utiliza el nombre fuente. La opción `-C toolconfs/hinit.conf` se utiliza para especificar el nombre del archivo de configuración utilizado con esta herramienta. La opción `-D` provoca que el conjunto de parámetros de la configuración actual se despliegue antes y después de ejecutar la herramienta. La opción `-M hmms/hmm.0` almacena los archivos del modelo macro HMM en el directorio especificado. A través de `-T 1` se activa la bandera de rastreo (Trace) para reportar el progreso básico.

“proto/NINE” es el nombre del archivo que contendrá el proptotipo HMM y las trayectorias siguientes especifican los archivos de contienen los ejemplos de entrenamiento.

Llamada de la herramienta HRest para el HMM ONE

```
HRest -A -u tmvw -w 3 -v 0.05 -i 10 -L labels/bcplabs/mon -l  
ONE -C toolconfs/hrest.conf -D -M hmms/hmm.1 -T 1  
hmms/hmm.0/ONE data/train/tr1.mfc data/train/tr10.mfc  
data/train/tr11.mfc data/train/tr12.mfc data/train/tr13.mfc  
data/train/tr14.mfc data/train/tr15.mfc data/train/tr2.mfc  
data/train/tr3.mfc data/train/tr4.mfc data/train/tr5.mfc  
data/train/tr6.mfc data/train/tr7.mfc data/train/tr8.mfc  
data/train/tr9.mfc
```

Llamada de la herramienta HRest para el HMM NINE

```
HRest -A -u tmvw -w 3 -v 0.05 -i 10 -L labels/bcplabs/mon -l  
NINE -C toolconfs/hrest.conf -D -M hmms/hmm.1 -T 1  
hmms/hmm.0/NINE data/train/tr1.mfc data/train/tr10.mfc  
data/train/tr11.mfc data/train/tr12.mfc data/train/tr13.mfc  
data/train/tr14.mfc data/train/tr15.mfc data/train/tr2.mfc  
data/train/tr3.mfc data/train/tr4.mfc data/train/tr5.mfc  
data/train/tr6.mfc data/train/tr7.mfc data/train/tr8.mfc  
data/train/tr9.mfc
```

La herramienta HRest se utiliza directamente a los modelos generados por Hlnit y considera que tiene como entrada la definición de los HMM inicializados, y utiliza la reestimación de Baum-Welch en lugar del entrenamiento de Viterbi. Esto involucra encontrar la probabilidad de estar en cada estado en cada tiempo de la trama utilizando el

algoritmo de Forward-Backward. Esta probabilidad se utiliza para formar los promedios ponderados para los parámetros de los HMM.

En la herramienta HRest la opción `-A` provoca que los argumentos de la línea de comando se impriman. Con HRest se actualizan todos los parámetros de los HMM, estos son, la media (m), la varianza (v), ponderaciones mezcladas (mixture weights—“ w ”) y probabilidades de transición (t), con la opción `-u tmvw` provoca que los parámetros indicados por los argumentos de las banderas se actualicen. La opción `-w 3` provoca que cualquier probabilidad de una observación ponderada mezclada o discreta que se encuentre por debajo de la constante global MINMIX se considere como cero. La opción `-v 0.05` establece la varianza mínima para el valor real indicado. La opción `-i 10` establece el número máximo de ciclos de reestimación, en caso de omitir la opción se utiliza por default 20. La opción `-L labels/bcplabs/mon`, establece la trayectoria donde se encuentran los archivos etiquetados, por default busca en el directorio actual. La opción `-l NINE` establece el nombre de la etiqueta del segmento, HRest busca en todos los archivos de entrenamiento y selecciona todos los segmentos con la etiqueta establecida. Cuando no se establece esta opción se considera que el archivo tiene un solo símbolo. La opción `-C toolconfs/hrest.conf` se utiliza para especificar el nombre del archivo de configuración utilizado con esta herramienta. La opción `-D` provoca que el conjunto de parámetros de la configuración actual se despliegue antes y después de ejecutar la herramienta. La opción `-M hmms/hmm.1` almacena los archivos del modelo macro HMM en el directorio especificado. A través de `-T 1` se activa la bandera de rastreo (Trace) para reportar el progreso básico. “hmms/hmm.0/NINE” es el nombre del archivo que contendrá el prototipo HMM que será re-estimado y las trayectorias siguientes especifican la lista de los archivos de entrenamiento.

Re-estimación incrustada (embedded)

```
HERest -A -w 3 -v 0.05 -C toolconfs/herest.conf -u tmvw -d
hmms/hmm.1 -D -M hmms/hmm.2 -L labels/bcplabs/mon -t 2000.0 -T
1 lists/bcplist data/train/tr1.mfc data/train/tr10.mfc
data/train/tr11.mfc data/train/tr12.mfc data/train/tr13.mfc
data/train/tr14.mfc data/train/tr15.mfc data/train/tr2.mfc
data/train/tr3.mfc data/train/tr4.mfc data/train/tr5.mfc
data/train/tr6.mfc data/train/tr7.mfc data/train/tr8.mfc
data/train/tr9.mfc
```

La fase de entrenamiento principal involucra el uso de la herramienta HREst la cual realiza entrenamiento incrustado (embedded), el cual utiliza también el Procedimiento de Baum-Welch utilizado para los casos aislados pero en lugar de entrenar cada modelo individualmente, todos los modelos se entrenan en paralelo.

La herramienta HREst carga todos los modelos de hmms/tmp (indicados por la opción -d) los cuales se encuentran indicados en la lista de modelos en labels/bcplabs/mon. Estos son reestimados utilizando la lista de archivos de entrenamiento mfc y el nuevo modelo obtenido se almacena en el directorio hmms/hmm.2 (especificado por la opción -H). La opción -A, -w, -v, -C y -T son similares a las de la herramienta HRest

Evaluación del Reconocimiento

En este momento el reconocedor se encuentra completo, ya se tiene construido el diccionario y la red de reconocimiento. Primero se realiza un reconocimiento con los mismos datos de entrenamiento.

Reconocimiento con conjunto de archivos de entrenamiento.

```
HVite -A -C toolconfs/hvite.conf -d hmms/hmm.2 -l test -w
networks/monLattice -D -L test -t 300.0 -T 1 lists/bcpvocab
lists/bcplist data/train/tr1.mfc data/train/tr10.mfc
data/train/tr11.mfc data/train/tr12.mfc data/train/tr13.mfc
data/train/tr14.mfc data/train/tr15.mfc data/train/tr2.mfc
data/train/tr3.mfc data/train/tr4.mfc data/train/tr5.mfc
data/train/tr6.mfc data/train/tr7.mfc data/train/tr8.mfc
data/train/tr9.mfc
```

En el directorio data/train se tienen los archivos de entrenamiento codificados, cada uno de ellos será reconocido y su transcripción de salida se colocará en el directorio test (especificado por la opción -l).

HVite cargará el archivo de red (indicado por la opción -w networks/monLattice) y conjuntará ésta con cada uno de los archivos de prueba (en este caso de entrenamiento). En lists/bcpvocab lists/bcplist se tiene la lista de modelos requeridos para la red. La opción -d hmms/hmm.2 especifica el directorio en el que se buscará la definición de los HMMs correspondientes a las etiquetas utilizadas en la red de reconocimiento. La opción -L especifica el directorio para encontrar la etiqueta de entrada. La opción -D provoca que el conjunto de parámetros de la configuración actual se despliegue antes y

después de ejecutar el comando. La opción `-t 300.0` habilita la transmisión de búsqueda que caiga f por debajo del máximo de todos los modelos se desactiva. Colocando f en 0.0 deshabilita este mecanismo.

Resultados del Reconocedor

```
HResults -A -L labels/bcplabs/mon lists/bcplist test/tr1.rec
test/tr10.rec test/tr11.rec test/tr12.rec test/tr13.rec
test/tr14.rec test/tr15.rec test/tr2.rec test/tr3.rec
test/tr4.rec test/tr5.rec test/tr6.rec test/tr7.rec
test/tr8.rec test/tr9.rec
```

Una vez que los datos de prueba han sido procesados por el reconocedor, el siguiente paso es analizar los resultados. Lo cual se hace a través de la herramienta HResults, la cual compara las transcripciones de salida de HVite con la transcripción de referencia original y genera varias estadísticas.

El trabajo de conjuntar de manera óptima las cadenas se realiza calculando un puntaje para esta comparación con respecto a la referencia de tal forma que etiquetas idénticas tienen una puntuación de 0, los errores de inserciones (I) y las omisiones (D) generan un puntaje de 7 y las sustituciones (S) un puntaje de 10². El porcentaje correcto (%Corr) del reconocedor se calcula con la siguiente expresión:

$$Porcentaje_Correcto = \frac{N - D - S}{N} \times 100\% \dots\dots\dots(9)$$

Donde N es el número total de etiquetas en la transcripción de referencia.

Para algunos propósitos el porcentaje de precisión (Acc) en el que se incluyen los errores de inserción, se define como:

$$Porcentaje_de_Precisión = \frac{N - D - S - I}{N} \times 100\% \dots\dots\dots(10)$$

HResults -A -L labels/bcplabs/mon lists/bcplist test
 En el commando HResult utilizado en el reconocedor de dígitos, se especifican los archivos conteniendo la transcripción para el reconocimiento (`test/tr1.rec ...`). La opción `-L` especifica el directorio donde se buscarán los archivos etiquetados. En la

trayectoria /bcplabs/mon se especifica la lista de todos los nombres de los modelos. Las estadísticas de reconocimiento, se pueden generar individualmente para cada archivo utilizando la opción `-f` y se puede generar una matriz de confusión si agregamos la opción `-p`.

Prueba con conjunto de archivos de prueba.

Una vez realizado el reconocimiento con los mismos datos de entrenamiento, se realiza el reconocimiento con los datos de prueba.

```
HVite -A -C toolconfs/hvite.conf -d hmms/hmm.2 -l test -w  
networks/monLattice -D -L test -t 300.0 -T 1 -p 5.0 -s 0.0  
lists/bcpvocab lists/bcplist data/test/te1.mfc  
data/test/te2.mfc data/test/te3.mfc
```

```
HResults -A -s -p -L labels/bcplabs/mon lists/bcplist  
test/te1.rec test/te2.rec test/te3.rec
```

Para cada Reconocedor se muestran los resultados obtenidos a través de la matriz de confusión. En la parte superior de cada matriz (a manera de columnas) se muestran los modelos de los dígitos generados por los modelos ocultos de Markov, mientras que en cada renglón se indica, para cada uno de los dígitos con qué modelo fueron reconocidos.

1) El primer Reconocedor de Dígitos fue entrenado con nativos (siete hablantes del idioma inglés) y se probó también con nativos (tres hablantes nativos del idioma inglés), diferentes a los utilizados para el entrenamiento. Como resultado se obtuvo un reconocimiento del 98.43%.

```

----- Overall Results -----
SENT: %Correct=0.00 [H=0, S=3, N=3]
WORD: %Corr=98.48, Acc=-1607.58 [H=65, D=0, S=1, I=1126, N=66]
----- Confusion Matrix -----
      S  O  T  T  F  F  S  S  E  N  T  Z
      I  N  W  H  O  I  I  E  I  I  E  E
      L  E  O  R  U  V  X  V  G  N  N  R
          E  R  E      E  H  E      O
          E                      N  T          Del [ %c / %e]
SIL  33  0  0  0  0  0  0  0  0  0  0  0
ONE  0  3  0  0  0  0  0  0  0  0  0  0
TWO  0  0  3  0  0  0  0  0  0  0  0  0
THRE 0  0  0  3  0  0  0  0  0  0  0  0
FOUR 0  0  0  0  3  0  0  0  0  0  0  0
FIVE 0  0  0  0  0  3  0  0  0  0  0  0
SIX  0  0  0  0  0  0  3  0  0  0  0  0
SEVE 0  0  0  0  0  0  0  3  0  0  0  0
EIGH 0  0  0  0  0  0  0  0  3  0  0  0
NINE 0  0  0  0  0  0  0  1  0  2  0  0 [66.7/1.5]
TEN  0  0  0  0  0  0  0  0  0  0  3  0
ZERO 0  0  0  0  0  0  0  0  0  0  0  3
Ins  491 15  8 27 126 140 70 91 40 41 3 74
=====
    
```

Tabla 1 Reconocedor entrenado con nativos y probado con nativos

2) Se realizó un segundo experimento, en el que se tomaron los modelos entrenados con nativos (siete hablantes del idioma inglés) pero esta vez se probó con hablantes no nativos (cinco hablantes hispanos). Para este segundo caso se obtuvo un porcentaje de reconocimiento del 79.74%. Podemos notar que el porcentaje de reconocimiento es menor al anterior debido a la pronunciación de los hispanos en el idioma inglés.

```

----- Overall Results -----
SENT:%Correct=0.00 [H=0, S=5, N=5]
WORD:%Corr=79.74,Acc=-10.34 [H=185,D=2, S=45, I=209, N=232]
----- Confusion Matrix -----
      S O T T F F S S E N T Z
      I N W H O I I E I I E E
      L E O R U V X V G N N R
          E R E     E H E     O
          E           N T
SIL 119  0  0  0  0  0  0  0  0  0  0  0  0
ONE   6  1  0  0  0  1  0  1  0  1  0  0  0 [10.0/3.9]
TWO   4  3  1  1  0  0  0  0  0  0  0  0  1 [11.1/3.4]
THRE  0  0  0  9  0  0  0  1  0  0  0  0  0 [90.0/0.4]
FOUR  0  0  0  0 10  0  0  0  0  0  0  0  0
FIVE  1  0  1  0  0  7  0  0  1  0  0  0  0 [70.0/1.3]
SIX   1  0  0  0  0  0  9  0  0  0  0  0  0 [90.0/0.4]
SEVE  0  0  0  1  0  0  0  9  0  0  0  0  0 [90.0/0.4]
EIGH  3  0  0  0  0  0  1  0  5  0  0  0  0 [55.6/1.7]
NINE  2  0  0  0  0  0  0  1  0  9  0  0  0 [75.0/1.3]
TEN   3  0  0  0  0  0  0  1  0  0  5  1  0 [50.0/2.2]
ZERO  3  0  5  0  0  0  0  1  0  1  0  1  1 [ 9.1/4.3]
Ins  108 3  1 12  8 12 16 17  4 28  0  0
    
```

=====

Tabla 2 reconocedor entrenado con nativos y probado con no nativos

3) Posteriormente se realizó otro reconocedor, pero esta vez estuvo entrenado (con siete hablantes) y probado con no nativos hablando inglés (tres hablantes hispanos), los archivos de prueba correspondieron a personas diferentes a las utilizadas para el entrenamiento. Se obtuvo el siguiente porcentaje de reconocimiento: 93.94%

```

----- Overall Results -----
SENT: %Correct=0.00 [H=0, S=3, N=3]
WORD: %Corr=93.94, Acc=-1716.67 [H=62, D=0, S=4, I=1195, N=66]
----- Confusion Matrix -----
      S O T T F F S S E N T Z
      I N W H O I I E I I E E
      L E O R U V X V G N N R
          E R E E H E O
          N T
SIL 33 0 0 0 0 0 0 0 0 0 0 0 0 Del [ %c / %e]
ONE 0 3 0 0 0 0 0 0 0 0 0 0 0
TWO 0 0 3 0 0 0 0 0 0 0 0 0 0
THRE 0 0 0 3 0 0 0 0 0 0 0 0 0
FOUR 0 0 0 0 3 0 0 0 0 0 0 0 0
FIVE 0 0 0 0 0 3 0 0 0 0 0 0 0
SIX 0 0 0 0 0 0 2 1 0 0 0 0 0 [66.7/1.5]
SEVE 0 0 0 0 0 0 0 3 0 0 0 0 0
EIGH 0 0 0 0 0 0 0 0 2 0 0 1 0 [66.7/1.5]
NINE 0 0 0 0 0 0 0 0 0 3 0 0 0
TEN 0 0 0 0 0 0 0 0 0 0 1 2 0 [33.3/3.0]
ZERO 0 0 0 0 0 0 0 0 0 0 0 3 0
Ins 417 69 13 55 70 43 36 16 47 32 0 397
=====
    
```

Tabla 3 reconocedor entrenado con no nativos y probado con no nativos

Se puede observar que cuando se prueba con las características de hablantes similares a los que se entrenó, el porcentaje de reconocimiento es alto, sin embargo cuando se prueba con un grupo de hablantes con diferentes características (en este caso pronunciando un idioma no nativo) el porcentaje de reconocimiento es menor.

Capítulo II: Técnicas de Adaptación

Técnicas de Adaptación

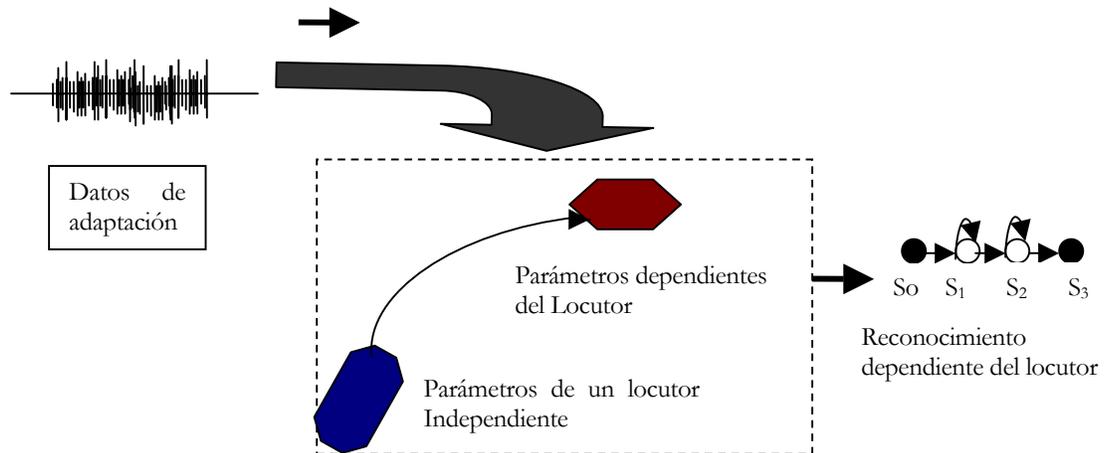


Fig.11 Proceso de Adaptación de un locutor

La idea fundamental en la adaptación de un nuevo locutor, consiste en un proceso en el cual se utilizan los datos de adaptación del nuevo locutor para ajustar el proceso modelado, de tal forma que los modelos representen un modelo más cercano a los datos de adaptación, como se representa en la figura 11. Básicamente se desea utilizar una cantidad de datos de adaptación tan pequeña como sea posible.

Las técnicas de adaptación de locutores para sistemas de reconocimiento de voz basadas en HMM se encuentran en dos categorías. La primera que emplea métodos que transforman la señal de voz de entrada del nuevo locutor en un vector (vector space) que es común con la señal de voz de entrenamiento, lo cual se conoce como técnica de mapeo espectral. Un segundo método transforma los parámetros del modelo para lograr una mejor combinación de las

características de los datos de adaptación, esta técnica se conoce como Método de mapeo de Modelo.

2.1 Método de Mapeo Espectral

Este método se basa en la consideración de que un sistema de reconocimiento puede mejorarse al emparar las características de los vectores de un nuevo locutor con los vectores de los datos de entrenamiento. El mapeo es diseñado de tal forma que las diferencias entre el conjunto de vectores de referencia y el conjunto de vectores mapeados sean minimizadas. Estas diferencias se deben a las diferencias espectrales de los sistemas producto de las señales de voz de los hablantes (por ejemplo, la forma y longitud del tracto vocal).

Se tiene referencia de algunos intentos utilizando esta técnica, en la que emplean planillas (templates) espectrales para ajustar los sistemas, generando una transformación para minimizar la diferencia entre el nuevo locutor y el hablante de referencia, como se muestra en la figura 12. Otras investigaciones han mapeado las características de los vectores tanto de los datos de entrenamiento como del nuevo hablante en un espacio de vectores común el cual maximiza la correlación entre los dos conjuntos.

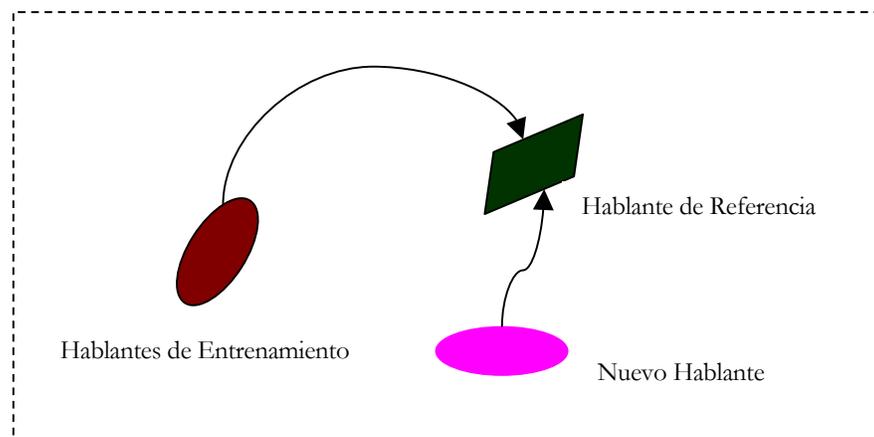


Fig. 12 Método de Mapeo Espectral.

2.2 Técnica de Mapeo de Modelado.

El objetivo del mapeo espectral, consiste en mejorar el acercamiento entre el hablante de referencia y el nuevo hablante. Sin embargo, este propósito no trata explícitamente de mejorar la precisión de los modelos para el nuevo hablante y, no tiene la ventaja completa de la adaptación de datos. Esto es lo que la técnica de mapeo de modelo trata de mejorar. En lugar de tratar de mapear a todos los hablantes a un espacio, ajusta los parámetros del modelo para representar mejor al nuevo hablante.

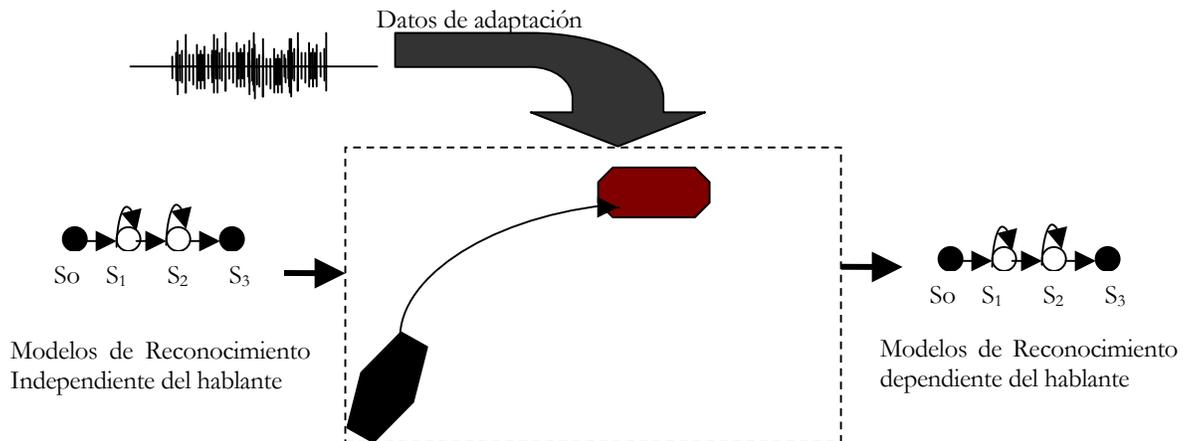


Fig. 13 Adaptación de un hablante por la técnica de mapeo de modelado

Un ejemplo común del método de mapeo de modelado es la técnica bayesiana MAP (maximum a posteriori) para adaptar Modelos Ocultos de Markov. En la técnica MAP la transformación se escoge de tal forma que los nuevos parámetros del modelo maximicen la función de probabilidad.

$$P(\lambda / O) = \frac{P(O / \lambda)P(\lambda)}{P(O)} \dots\dots\dots(11)$$

Donde O es la secuencia de observaciones de adaptación y λ es el conjunto de parámetros que definen la distribución. Se han utilizado

diferentes métodos para estimar el valor de λ incluyendo una técnica de k-media segmentada (segmental k-means approach) y un método basado en EM.

La mayoría de métodos MAP están limitados debido a que solo adaptan los parámetros que son directamente observados en la adaptación de datos.

2.3 Técnica de Adaptación MLLR

Una técnica de mapeo de modelo más exitosa es Maximum likelihood linear regresión (MLLR). MLLR fue diseñado para reducir las desventajas de las técnicas de mapeo de espectro y mapeo de modelo. MLLR es un método basado en transformadas, el cual adapta los parámetros del modelo y solo requiere una pequeña cantidad de datos de entrenamiento.

Maximum Likelihood Linear Regression (MLLR) es una técnica de adaptación que estima un conjunto de transformaciones lineales para los parámetros de la media y la varianza de las Gaussianas en un sistema HMM. El efecto de estas transformaciones consiste en mover la media y alterar la varianza en el sistema inicial de tal forma que cada estado en el sistema HMM sea más probable de generar la adaptación. De esta forma la adaptación a nuevas personas puede realizarse utilizando una pequeña cantidad de datos de entrenamiento o adaptación. La matriz de transformación usada para dar una nueva estimación de la media adaptada está dada por:

$$\hat{\mu} = W\xi, \dots\dots\dots(12)$$

donde W es la matriz de transformación n x (n+1), n es la dimensionalidad de los datos y ξ es el vector de la media extendido,

$$\xi = [\omega \mu_1 \mu_2 \dots \mu_n]^T \dots\dots\dots(13)$$

Donde ω representa un 'offset' cuyo valor está fijo en 1.

W se puede descomponer en:

$$W=[b A], \dots\dots\dots(14)$$

Donde A representa la matriz de transformación n x n y b representa el vector desplazado. La matriz de transformación W se obtiene

resolviendo un problema de maximización usando la técnica de “Expectation-Maximization” (EM), la cual se utiliza para calcular la matriz de transformación de la varianza.

La adaptación MLLR se puede aplicar de una manera flexible, dependiendo de la cantidad de datos de adaptación que se tengan disponibles. Si se cuenta con una cantidad reducida de datos, entonces se puede realizar una transformación global de la adaptación. La transformación global se aplica a cada componente de la Gaussiana de los modelos. Pero a medida que se tiene una cantidad mayor de datos de adaptación, se puede mejorar la adaptación incrementando el número de transformaciones. Cada transformación es más específica y se aplica a cierto grupo de componentes de la Gaussiana

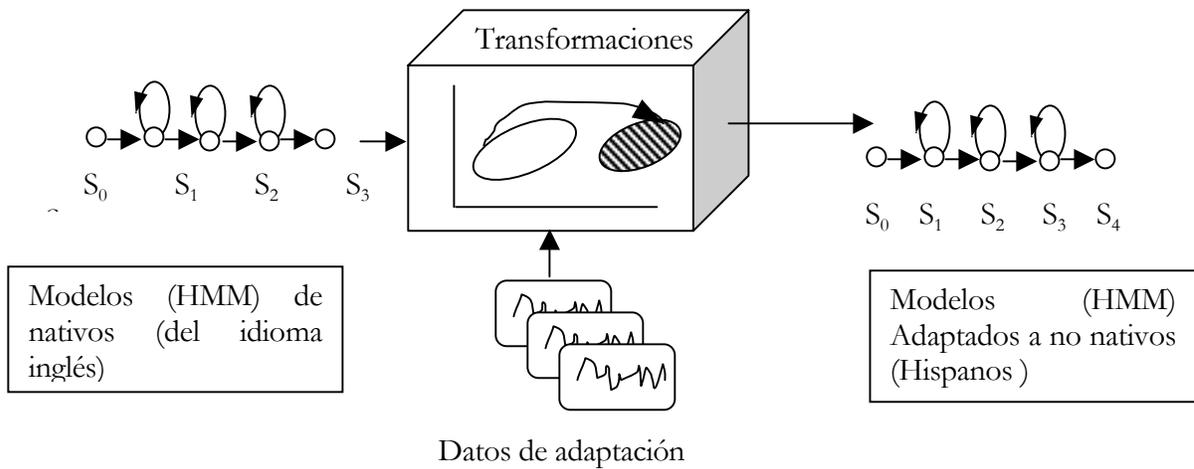


Fig. 14 Adaptación de HMMs

Capítulo III: Reconocedor de Voz Adaptado

3.1 Adaptación de los Modelos Ocultos de Markov.

Como primera fase del proyecto se generaron Modelos Ocultos de Markov a partir de la base de datos TIMIT, de tal manera que el reconocimiento de hispanos hablando inglés es por consecuencia bajo, debido a la pronunciación y tono de voz entre otras características. Para superar estas limitaciones, se tienen básicamente dos opciones: construir un conjunto de modelos con hablantes independientes (hispanos hablando inglés), lo cual requeriría una gran cantidad de datos de entrenamiento procedente de una variedad de personas. Otra alternativa consiste en adaptar los modelos actuales obtenidos (modelos nativos del inglés) a las características de los nuevos hablantes (hispanos) usando solo una pequeña cantidad de datos de adaptación o entrenamiento. Las técnicas de adaptación se aplican a un conjunto de modelos independientes bien entrenados de locutores con el fin de adaptarlo para que adquiera las características de un hablante particular.

Las técnicas de adaptación del nuevo hablante pueden realizarse de diferentes modos. Si la transcripción de los datos de adaptación es conocida, entonces se llama adaptación supervisada, si los datos de adaptación no son conocidos entonces se llama adaptación no supervisada. En el caso donde todos los datos de adaptación están disponibles (registrados) entonces se llama adaptación estática. Alternativamente la adaptación se puede realizar incrementalmente a medida que los datos de la adaptación son disponibles, lo cual se conoce como adaptación incremental. En este proyecto nos enfocaremos a la transcripción estática supervisada.

3.2 Desarrollo del Reconocedor de Voz Adaptado.

Para la realización de los Reconocedores de voz adaptados se utilizaron dos bases de datos, la de locutores nativos del inglés, que consta de frases cortas; y la no nativa (locutores hispanos hablando inglés) pronunciando las mismas frases cortas.

Los datos de voz son necesarios para el entrenamiento y prueba de los modelos que se realizarán. Los datos necesarios para el entrenamiento se tomaron de la base de datos fonética-acústica TIMIT. Esta información se utiliza junto con un diccionario de pronunciación que proporciona el nivel de transcripción inicial que permitirá iniciar el proceso de entrenamiento de los Modelos Ocultos de Markov.

Preparación de datos

Al igual que en el desarrollo del reconocedor de dígitos, el primer paso para desarrollar la adaptación de un reconocedor, consiste en la preparación de datos. En este caso se requieren de los datos de voz de los nuevos hablantes con los que se adaptará el modelo y se probará el sistema adaptado.

La cantidad de datos de adaptación se recomienda que sea de por lo menos 30 segundos de señal de voz, lo cual corresponde aproximadamente a 20 expresiones.

A través de la herramienta HTK se proporciona la gramática para el reconocedor que se requiere. Para esto se cuenta con un lenguaje de definición de gramática, el cual consiste de un conjunto de definiciones de variables y posteriormente se proporciona una expresión en la que se describen las palabras que deben ser reconocidas. En este tipo de definición de gramática, cada barra vertical denota las diferentes alternativas de palabras que pueden existir, los paréntesis cuadrados denotan elementos opcionales y los paréntesis normales denotan una o más repeticiones.

Para este Reconocedor, la definición de todas las palabras posibles que se podrían pronunciar en las diferentes frases, se realizó a través de la variable \$palabras, y estas palabras se pueden estar repitiendo para formar las frases.

A continuación se muestra la gramática empleada en el Reconocedor de frases, para crear la red.

```
$palabras = sil | yesterday | i | drove | to | my | brothers |  
house | and | had | lunch | with | he | and | his | family | in  
| miami | it | is | usually | warm | and | never | snows | have  
| to | work | today | but | tomorrow | my | day | off | they |  
watched | the | ship | named | constitution | as | she | slipped  
| below | waves | kansas | states | new | basketball | tickets |  
cost | ten | dollars | now | cowboys | ride | big | bulls |  
horses | couple | ran | on | looking | sandy | beach | for |  
shells | cabbage | patch | kids | are | adults | that | adopt |  
little | girls | their | mothers | baby | play | house | but |  
boys | trains | we | relaxed | passed | through | tiny |  
mexican | shops;
```

```
(SENT-START <$palabras> SENT-END)
```

En este caso la gramática queda muy simple y se deja en libertad de repetir las palabras, para ser menos restrictivo en la pronunciación de las oraciones.

Inicialmente, se intentó otra gramática diferente para este reconocedor, la cual se muestra a continuación.

```
$palabra1 = yesterday | cowboys | the | baby;  
$palabra2 = i | ride | couple | girls;  
$palabra3 = drove | big | ran | play;  
$palabra4 = to | bulls | on | house;  
$palabra5 = my | and | the | but;  
$palabra6 = brothers | horses | sandy | baby;  
$palabra7 = house | beach | boys | and | looking | play | had |  
for | with | lunch | shells | trains | with | he | his | family  
| shell;
```

```
(SENT-START ([sil] $palabra1 [sil] $palabra2 [sil] $palabra3  
[sil] $palabra4 [sil] $palabra5 [sil] $palabra6 [sil]  
<$palabra7 [sil]>)SENT-END)
```

Este tipo de gramática mostró ser más restrictiva en cuanto a que establece una secuencia en la que se producirán las palabras dentro de las frases y permite reducir el número de opciones que puede encontrar lo cual dependiendo de la aplicación, puede mejorar los resultados. En aplicaciones en los que únicamente son palabras sin secuencia se recomienda la primer gramática mencionada.

Una recomendación que se puede tener en un sistema orientado en una secuencia de palabras establecidas por oraciones, es que, preferentemente las oraciones tengan la misma longitud de palabras, aunque se puede tener situaciones en las que los hablantes al darse cuenta que no pronuncian adecuadamente una palabra, tiendan a repetir la palabra sin haber terminado la oración, o repitan nuevamente la oración, por lo que los resultados del reconocedor pueden variar por la gramática establecida.

La red creada para estas gramáticas, la obtenemos a través de la herramienta

```
Hparse gram_frases wdenet_frases
```

La gramática es importante, ya que genera la red de palabras, la cual se define utilizando una notación de bajo nivel que en HTK se conoce como Formato de Entramado Estándar (Estándar lattice Format, "SLF") en la cual cada palabra referenciada, o cada transición de palabra a palabra está listada explícitamente. La herramienta anterior crea una red de palabras equivalente a la gramática establecida, esta red de palabras se almacena en el archivo `wdenet_frases`.

Una vez que se tiene la gramática y la red, se debe crear un diccionario, para lo cual primero se debe crear una lista de palabras requeridas en las oraciones que se utilizarán como datos de entrenamiento. Para la construcción de modelos acústicos robustos, se necesita que el entrenamiento incluya un conjunto grande de oraciones con muchas palabras. Para el caso del reconocedor de frases, la lista de palabras de entrenamiento utilizada, fue la siguiente (archivo `wlist`):

Adopt	now
adults	off
and	on
are	passed
as	patch
baby	play
baseball	ran
basketball	relaxed
beach	ride
below	sandy
big	she
boys	shell
brother	shells
brother's	ship
bulls	slipped
but	shop
cabbage	shops
constitution	sil
cost	silence
couple	snows
cowboys	sp
day	state's
dollars	ten
dolls	that
drove	the
family	their
for	they
girls	through
had	tickets
have	tiny
he	to
him	today
his	tomorrow
horses	train
house	trains
i	usually
in	warm
is	watched
it	waves
kansas	we
kids	with
little	work
looking	yesterday
lunch	
mexican	
miami	
mothers	
my	
named	
never	
new	

Para crear el diccionario, se usa la herramienta HDMan, y necesitamos emplear un diccionario de pronunciación, por ejemplo el de Inglés Británico BEEP, o el de Inglés Americano de TIMIT. Es importante notar que la pronunciación de palabras (por fonio) en Inglés Británico e Inglés Americano tienen algunas variaciones, y de acuerdo a la orientación que se le quiera dar al reconocedor se deberá seleccionar el diccionario adecuado. Para los Reconocedores realizados, se utilizó el diccionario de pronunciación de Inglés Americano TIMIT (dic_timit). A través de la herramienta:

```
HDMan -m -w wlist -n monophones -l dlog dict dic_timit
```

De donde:

‘wlist’ es la lista de palabras que serán reconocidas y que aparecen en la gramática anterior,

‘monophones’ es un archivo de salida que contiene la lista de fonios usados.

‘dict’ es el nuevo diccionario creado al buscar las palabras fuentes en ‘dic_timit’ y encontrar la pronunciación de cada palabra en ‘wlist’.

La opción ‘-l’ se utiliza para obtener un archivo de salida ‘dlog’ que contiene estadísticas del diccionario construido. Indicando si hay palabras que faltan.

Una vez que los datos de entrenamiento y prueba se han grabado, se puede estimar un Modelo Oculto de Markov para cada uno de estos fonios.

A continuación se muestra la forma que tiene el archivo de pronunciación de palabras en Inglés Americano. Podemos observar que cada palabra contiene una secuencia de fonios. En los paréntesis cuadrados se muestra la cadena de salida obtenida cuando la palabra es reconocida. En el caso de que sean omitidos, la palabra misma será la salida. En caso de existir los paréntesis cuadrados, pero vacíos, entonces no se tendrá salida (tal es el caso de los símbolos SENT-START y SENT-END, cuya pronunciación se marca con un modelo de sil, y provoca una salida nula). También se puede tener palabras con múltiples pronunciaciones.

Fragmento del diccionario de pronunciación TIMIT

```

a                ax sp
abbreviate      ax b r iy v iy ey t sp
abdomen         ae b d ax m ax n sp
abides          ax b ay d z sp
ability         ax b ih l ix t iy sp
able            ey b el sp
ably            ey b l iy sp
abolish         ax b aa l ih sh sp
aborigine       ae b axr ih jh ix n iy sp
aborigines      ae b axr ih jh ix n iy z sp
about           ax b aw t sp
above           ax b ah v sp
abruptly        ax b r ah p t l iy sp
absences        ae b s ix n s ix z sp
absent          ae b s ix n t sp
.
.
back            b ae k sp
backed          b ae k t sp
backgrounds     b ae k g r aw n d z sp
backyards       b ae k y aa r d z sp
bacteria        b ae k t ih r iy ax sp
bad             b ae d sp
badness         b ae d n ix s sp
baffle         b ae f el sp
bag             b ae g sp
.
.
careful         k ae r f el sp
carefully       k ae r f uh l iy sp
carl            k aa r l sp

carol           k ae r el sp
carpet          k aa r p ax t sp
carriage        k ae r ih jh sp
carried         k ae r iy d sp
carries         k ae r iy z sp
carry           k ae r iy sp
carrying        k ae r iy ix ng sp
cars            k aa r z sp

SENT-END      [] sil
SENT-START    [] sil
    
```

Para los datos de Entrenamiento y Prueba, se ocuparon una serie de archivos grabados con personas hispanas del sexo masculino hablando en Inglés pronunciando frases cortas. A continuación se muestra un fragmento de la señal de voz de dos hablantes

pronunciando dos frases diferentes, estas señales se encuentran etiquetadas.

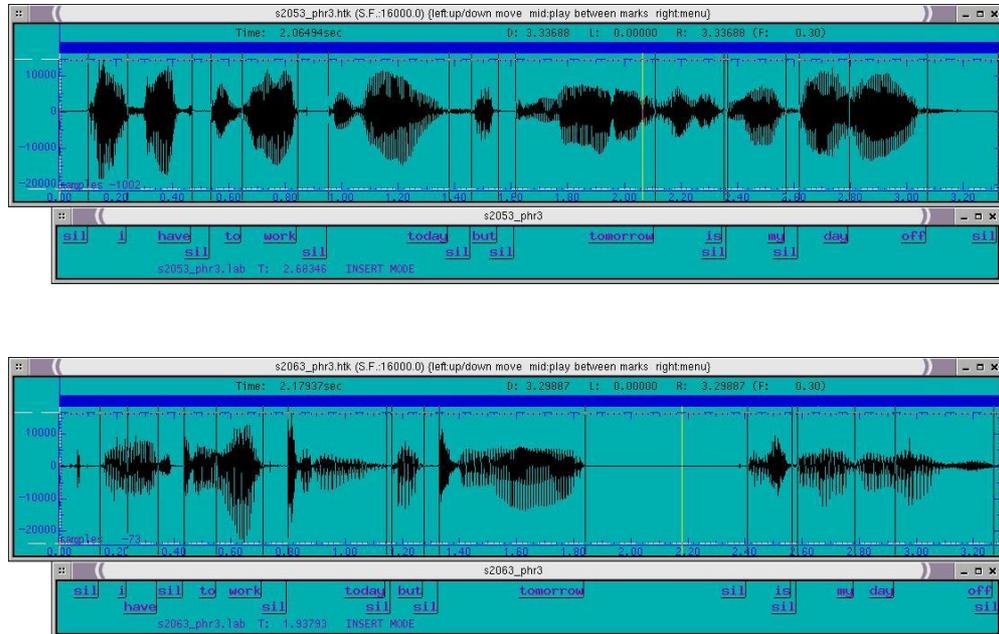


Fig. 15 Etiquetado de la Señal de Voz “frases” con xwaves.

Para realizar el etiquetado de todas las señales de voz que se utilizaron de los diferentes hablantes, se utilizó el software xwaves, una vez que se editaron las señales, se procedió a establecer las etiquetas de cada fragmento de la señal para separar por palabras las diferentes oraciones, para esto se estuvo escuchando cuidadosamente cada archivo grabado a través de las bocinas de la computadora y se obtuvo a través del software xwaves el espectro de la señal mostrando los formantes, esto ayudó a que se delimitara de manera más precisa las fronteras de cada palabra, los silencios y ruidos que presentaba la señal. El etiquetado se realiza con este software, ya que permite seleccionar fragmentos de la señal, reproducirlos y permite con su editor escribir la etiqueta que se asigna a ese fragmento de señal. En la siguiente figura podemos observar un ejemplo, para dos hablantes distintos, de este proceso de etiquetado, que como se mencionó se fue realizando en forma manual para cada archivo de voz de los diferentes locutores.

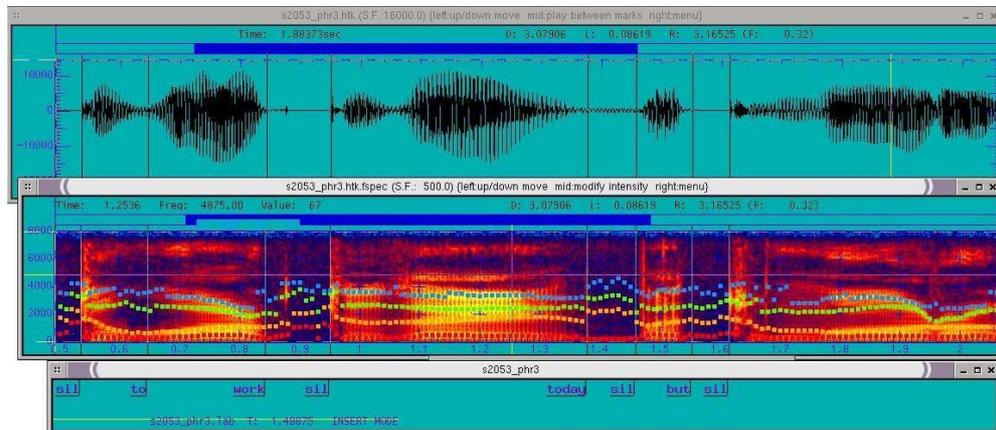


Fig. 16 Ejemplo “a” del etiquetado de la Señal de Voz “frases” con espectro.

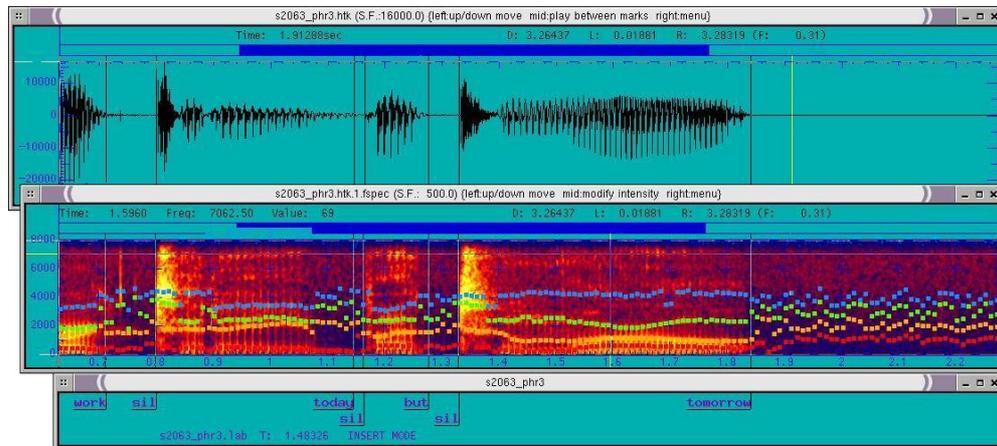


Fig. 17 Ejemplo “b” del etiquetado de la Señal de Voz “frases” con espectro.

Para realizar el grabado de un locutor en cualquier momento, se puede utilizar la herramienta HSLab del HTK, que a través de su menú, permite realizar el grabado (forma de onda) y es también una herramienta para etiquetar la señal.

A través de esta herramienta diferentes hablantes pueden grabar la pronunciación de las frases requeridas para el entrenamiento o prueba del reconocedor. Los archivos generados son con extensión .htk

Para el entrenamiento de los Modelos Ocultos de Markov, se debe asociar por cada archivo de entrenamiento una transcripción a nivel fonio. Primeramente es necesario que los archivos de señal se tengan etiquetados.

Para la transcripción es necesario tener un archivo con las palabras o frases pronunciadas de cada archivo y obtener la transcripción a nivel palabra. Esto es, obtener un archivo en el que se indique la trayectoria de cada archivo etiquetado y las palabras que contiene cada archivo, pero cada palabra debe ser escrita en un solo renglón y al final de la frase se debe terminar con un punto. Para realizar la transcripción a nivel palabra se utilizó el programa prompts2m1f.

A continuación se proporciona un ejemplo para uno de los hablantes.

```
s2015_phr1 yesterday i drove to my brother's house and had  
lunch with him and his family  
s2015_phr3 i have to work today but tomorrow is my day off
```

s2015_phr4 they watched the ship named the constitution as
she slipped below the waves
s2015_phr5 kansas state's new basketball tickets cost ten
dollars now
s2015_phr6 cowboys ride big bulls and horses
s2015_phr7 the couple ran on the sandy beach looking for the
shells
s2015_phr8 cabbage patch kids are adults that adopt little
girls as their mothers
s2015_phr9 baby girls play house but baby boys play with
trains

Fragmento del archivo arch_adaptaFV5.mlf

Utilizando el archivo "prompts2mlf", y dando como entrada el archivo anterior "arch_adaptaFV5.mlf" se obtiene como salida al archivo phrases_silFV5.mlf una vez ejecutado el comando siguiente.

```
Prompts2mlf_wsil phrases_silFV5sil.mlf arch_adaptaFV5.mlf
```

Un fragmento del archivo obtenido como resultado, tiene la siguiente forma:

```
#!MLF!#  
"/s2063_phr1.lab"  
yesterday  
sil  
i  
sil  
drove  
sil  
to  
sil  
my  
sil  
brother's  
sil  
house  
sil  
i  
sil  
drove  
sil  
to  
sil  
my  
sil  
brother's  
sil  
house  
sil
```

```
and
sil
had
sil
lunch
sil
with
sil
him
sil
and
sil
his
sil
family
sil
.
"/s2063_phr3.lab"
i
sil
have
sil
to
sil
work
sil
today
sil
but
```

Con ayuda de la herramienta HCopy se obtendrá el estado final de la preparación de datos, el cual consiste en parametrizar la formas de onda naturales del archivo de voz htk en una secuencia de vectores caracterizados, se utilizó coeficientes MFCC's ("Mel Frequency Cepstral Coefficients"). Se utilizó un archivo de configuración (config), en el que se establecen los parámetros de conversión.

```
# Coding parameters
TARGETKIND = MFCC_E
TARGETRATE = 100000.0
SAVECOMPRESSED = T
SAVEWITHCRC = T
WINDOWSIZE = 250000.0
USEHAMMING = T
PREEMCOEF = 0.97
NUMCHANS = 26
CEPLIFTER = 22
NUMCEPS = 12
ENORMALISE = V
```

Contenido del Programa de configuración config

En este archivo se especifica que los parámetros serán MFCC, el periodo de la trama es de 100 ms, utiliza una ventana Hamming y la señal debe tener un preenfasis de primer orden utilizando un coeficiente de 0.97. El banco de filtros debe tener 26 canales y 12 coeficientes MFCC de salida.

La parametrización se logra con la herramienta HCopy, de la siguiente manera:

```
HCOPY -C config -S adapt_FV5.scp  
HCOPY -C config -S testAdapt_FV5.scp
```

En el comando se especifican los archivos con los que se realizará la adaptación de los modelos y en adapt_FV5.scp los archivos de prueba.

Una vez que se tiene el archivo MLF a nivel palabra, el estado final involucra generar el contexto dependiente de la transcripción a nivel fonio de los datos de adaptación y la transcripción a nivel palabra de los datos de prueba. La transcripción de los datos de prueba puede obtenerse utilizando el archivo prompts2mlf como se indicó anteriormente. Para minimizar el problema de múltiples pronunciaciones la transcripción a nivel fonio de los datos de adaptación puede obtenerse utilizando la herramienta HVite, para realizar una alineación forzada de los datos de adaptación. Considerando que la transcripción a nivel palabra se encuentra en el archivo phrases_FV5sil.mlf, el siguiente comando colocará la transcripción a nivel fonio en adapt_FV5Phones.mlf.

```
HVite -l '*' -o SWT -C config_htk -a -H  
/home/ggma/hmm.9/macros -H /home/ggma/hmm.9/hmmdefs -i  
adapt_FV5Phones.mlf -m -t 250.0 -I phrases_FV5sil.mlf -y lab  
-S train_FV5.scp dict_tf monophones_t
```

En donde:

'-l dir' especifica el directorio donde se almacenan los archivos etiquetados de salida. Al omitirse se utilizará el mismo directorio.

'-o' selecciona cómo deben formatearse las etiquetas de salida,

SWT es una cadena que indica banderas binarias que controlan las opciones de formateo. "S" remueve la cuantificación de la etiqueta de salida. Por default la cuantificación se establece para indicar la probabilidad total del segmento. "W" No incluye las palabras en los archivos de etiqueta de salida cuando se realiza el modelo de alineación. "T" no incluye los tiempos en los archivos de etiquetas de salida.

'-C' se utiliza para especificar un archivo de configuración.

- '-a' realiza alineamiento. Hvite cargará un archivo de etiqueta y creará una red de alineamiento para cada archivo de prueba.
- '-H mmf' esta opción puede cargar múltiples MMF's
- '-i' Indica el nombre del archivo de transcripción de salida a MLF's
- '-m' el archivo de salida contendrá múltiples niveles. El primer nivel será el número de modelo y el segundo será el nombre de la palabra.
- '-t f' habilita la búsqueda guiada de cualquier modelo cuya máxima probabilidad caiga por debajo de f a partir del máximo para todos los modelos desactivándolos.
- '-l' Carga el archivo de etiquetado maestro mlf, esta opción puede repetirse para cargar diferentes MLF's
- '-y' coloca la extensión para los archivos de etiquetas de salida como ext (default rec)
- '-S' se utiliza para especificar un archivo Script

HVite cargará un archivo de red, empatará este con cada uno de los archivos de prueba y creará una nueva red para cada archivo de prueba del correspondiente archivo etiquetado (archivo .lab).

Fragmento del archivo phrases_FV5sil.mlf	Fragmento del archivo adapt_FV5Phones.mlf
<pre>#!MLF!# "/s2015_phr1.lab" yesterday sil i sil drove sil to sil my sil brothers sil house sil and sil had sil lunch sil with sil him sil</pre>	<pre>#!MLF!# "/s2015_phr1.lab" y eh s t axr d ey sp sil ay sp sil d r ow v sp sil t uw sp sil m ay</pre>

and	sp
sil	sil
his	b
sil	r
family	ah
sil	dh
.	axr
"*/s2015_phr3.lab"	z
i	sp
sil	sil
have	hh
sil	aw
to	s
sil	sp
work	sil
sil	ae
today	n
sil	d
but	sp
sil	sil
tomorrow	hh
sil	ae
is	d
sil	sp
my	sil
sil	l
day	ah
sil	n
off	ch
sil	sp
	sil
	w
	ih
	dh
	sp
	sil
	hh
	ih
	m
	sp
	sil
	ae
	n
	d
	sp
	sil
	hh
	ih
	z
	sp
	sil
	f
	ae
	m

	el iy sp sil . "/s2015_phr3.lab"
--	---

Generación de Transformadas

Dependiendo de la cantidad de datos de adaptación disponibles se realiza la adaptación, si solo se dispone de una pequeña cantidad de datos, se puede generar una transformada global para cada distribución de salida de cada modelo. A medida que se tienen mas datos de adaptación, se pueden generar transformadas más específicas para grupos específicos de Gaussianas. Para identificar el número de transformadas que se pueden estimar utilizando los datos de adaptación actuales, la herramienta HEAdapt utiliza un árbol de regresión de clase para agrupar las distribuciones de salida que sufrirán la misma transformación. La herramienta HHed se utiliza para construir un árbol de regresión de clase y almacenar parte del conjunto de HHM.

```
HHed -H /home/ggma/hmm.9/macros -H /home/ggma/hmm.9/hmmdefs  
-M /home/ggma/hmm.17 regtree.hed monophones_t
```

Esta herramienta crea un árbol de regresión de clase utilizando los modelos almacenados en hmm.9 (estos modelos previamente se generaron de manera similar a la indicada en el Reconocedor de Dígitos). Los modelos junto con la información del árbol de regresión clase se escribirán en hmm.17

El archivo regtree.hed contiene los siguientes comandos:

```
LS "Stats"  
RC 32 "rtree"  
RN native_models
```

"LS" carga el archivo de estadísticas de ocupación de estado generado por la última aplicación de HERest la cual creó los modelos en hmm.9 . "RC" trata de construir un árbol de regresión de clase con 32 nodos terminales o ramas utilizando las estadísticas,

La herramienta HEAdapt se utiliza par realizar la adaptación estática, donde todos los datos de adaptación se procesan en un simple bloque. Esta herramienta se compone de dos pasos. En el primer paso se realiza una adaptación global, en el segundo paso, se utiliza la transformación global para transformar el conjunto de modelos, dando lugar a una mejor alineación estado/trama, la cual posteriormente se utiliza para estimar un conjunto de transformadas específicas usando un árbol de regresión de clase. Una vez estimadas las transformadas la herramienta HEAdapt generará el nuevo conjunto de modelos adaptado en un archivo de modelo de transformadas (TMF). Los comandos utilizados para realizar la solicitud de las dos fases de la adaptación estática con HEAdapt se muestran a continuación:

```
HEAdapt -C config_htk -g -S train_2063_53.scp -I  
adapt_FV5Phones.mlf -H /home/ggma/hmm.17/macros -H  
/home/ggma/hmm.17/hmmdefs -K globalVF5.tmf monophones_t
```

```
HEAdapt -C config_htk -S train_2063_53.scp -I  
adapt_FV5Phones.mlf -H /home/ggma/hmm.17/macros -H  
/home/ggma/hmm.17/hmmdefs -J globalVF5.tmf -K rc_FV5  
monophones_t
```

Las transformaciones de regresión de clase y global se almacenan en el archivo globalVF5.tmf y rc_FV5.

Evaluación del Sistema Adaptado

Para evaluar el desempeño del sistema de adaptación, los datos de prueba que previamente se grabaron deben ser reconocidos con la herramienta HVite.

```
HVite -C config_htk -H /home/ggma/hmm.17/macros -H  
/home/ggma/hmm.17/hmmdefs -S testAdapt_FFV2.scp -l '*' -J  
rc_FV5 -i recoutAdapt_FV5.mlf -w wnet_frases -p 0.0 -s 5.0  
dict_tf monophones_t
```

El archivo testAdapt_FFV2.scp contiene la lista de los archivos de prueba codificados, la opción -J que se presenta en esta ocasión se utiliza para cargar el archivo de modelo de transformaciones rc_FV5. El resultado del conjunto de modelos adaptados se pueden obtener utilizando la herramienta HResults, como se indicó en el Reconocedor de Dígitos. La mejora obtenida en el modelo adaptado con respecto al modelo sin adaptar, se puede evaluar si se realiza el reconocimiento de los datos de prueba con el conjunto de modelos sin adaptar y comparando con los resultados del reconocedor con el conjunto de modelos adaptado.

```
HResults -p -f -t -I /home/ggma/corpora/phrasestest_FV2.mlf  
wlist_min recoutAdapt_FV5.mlf > logdic04_03VF5
```

Esta herramienta se aplica al cada archivo de reconocimiento. El archivo "wlist_min" contiene una lista de todos los nombres de los modelos. Por cada archivo de reconocimiento, un archivo de transcripción con el mismo nombre pero con extensión .lab se lee y se compara. La opción -p genera como salida una matriz de confusión, la opción -f acumula estadísticas para todos los archivos de entrada y proporciona un resumen a la salida. La opción -t provoca una transcripción de alineamiento en tiempo para cada archivo de prueba. La opción -I carga el archivo de etiquetas maestro (MLF) con la trayectoria especificada.

3.3 Resultados de los Experimentos Realizados

De investigaciones anteriores en las que se realizaron reconocedores de dígitos sin adaptar y adaptados [7] se tienen los siguientes resultados.

Tipo de Reconocedor	Porcentaje de Reconocimiento
Reconocedor de dígitos entrenado con personas nativas del Inglés y probado con nativos	Reconocimiento = 98.48%
Reconocedor de dígitos entrenado con personas nativas del Inglés y probado con no nativos (Hispanos)	Reconocimiento = 79.74%
Reconocedor de dígitos adaptado, entrenado con personas nativas del Inglés, adaptado con no nativos (Hispanos) y probado con no nativos.	Reconocimiento = 93.94 %
Reconocedor de dígitos entrenado con personas nativas del Inglés y probado con un solo hablante no nativo (Hispano)	Reconocimiento = 94.37%
Reconocedor de dígitos entrenado con personas nativas del idioma Inglés, adaptado con muestras de un Hispano y probado con el mismo Hispano .	Reconocimiento = 100%

Tabla 4. Resultados de los Reconocedores de dígitos

A continuación se muestran las matrices de confusión para el Reconocedor de dígitos entrenado con personas nativas del Inglés y probado **con un solo hablante no nativo (Hispano) y Reconocedor de dígitos entrenado con personas nativas del idioma Inglés, adaptado con muestras de un Hispano y probado con el mismo Hispano .**

```

===== HTK Results Analysis =====
Date: Wed May 15 20:11:09 2002
Ref : /home/ggma/Rene/data/16_all_sil.mlf
Rec : recout_nor.mlf
----- Overall Results -----
SENT: %Correct=0.00 [H=0, S=34, N=34]
WORD: %Corr=94.37, Acc=26.16 [H=1425, D=0, S=85, I=1030, N=1510]
----- Confusion Matrix -----
      z o t t f f s s e n t o o s
      e n w h o i i e i i e r u i
      r e o r u v x v g n n l
      o e r e e h e
      e n t
zero 79 0 0 0 0 0 0 0 0 0 0 0 1 0 Del [%c / %e]
one 0 8 17 1 0 0 0 0 14 16 1 0 0 11 0 [11.8/4.0]
two 0 0 66 0 0 0 0 0 0 0 0 0 0 0 0
thre 0 0 6 59 0 0 0 0 1 0 0 0 0 2 0 [86.8/0.6]
four 0 0 0 0 67 0 0 0 0 0 0 1 0 1 0 [97.1/0.1]
five 0 0 0 0 0 67 0 0 0 0 0 0 0 0 0
six 0 0 0 0 0 0 65 0 0 0 0 0 0 0 0
seve 0 0 0 0 0 0 0 70 0 0 0 0 0 0 0
eigh 0 0 4 0 0 0 0 0 61 0 0 0 0 2 0 [91.0/0.4]
nine 0 0 1 0 0 0 0 0 1 60 4 0 0 1 0 [89.6/0.5]
ten 0 0 0 0 0 0 0 0 0 0 68 0 0 0 0
sil 0 0 0 0 0 0 0 0 0 0 0 0 0 755 0
Ins 0 0 339 1 0 0 0 0 99 50 13 6 54 468
=====
    
```

Tabla 5 reconocedor entrenado con nativos del Inglés y probado con un hablante no nativo

```

===== HTK Results Analysis =====
Date: Fri Jun 21 19:29:48 2002
Ref : /home/ggma/Rene/data/16_all_sil.mlf
Rec : recoutAdapt.mlf
----- Overall Results -----
SENT: %Correct=0.00 [H=0, S=34, N=34]
WORD: %Corr=100.00, Acc=63.44 [H=1510, D=0, S=0, I=552, N=1510]
----- Confusion Matrix -----
      z o t t f f s s e n t o s
      e n w h o i i e i i e u i
      r e o r u v x v g n n l
      o e r e e h e
      e n t
zero 80 0 0 0 0 0 0 0 0 0 0 0 0 0 Del [%c/%e]
one 0 68 0 0 0 0 0 0 0 0 0 0 0 0
two 0 0 66 0 0 0 0 0 0 0 0 0 0 0
thre 0 0 0 68 0 0 0 0 0 0 0 0 0 0
four 0 0 0 0 69 0 0 0 0 0 0 0 0 0
five 0 0 0 0 0 67 0 0 0 0 0 0 0 0
six 0 0 0 0 0 0 65 0 0 0 0 0 0 0
seve 0 0 0 0 0 0 0 70 0 0 0 0 0 0
eigh 0 0 0 0 0 0 0 0 67 0 0 0 0 0
nine 0 0 0 0 0 0 0 0 0 67 0 0 0 0
ten 0 0 0 0 0 0 0 0 0 0 68 0 0 0
sil 0 0 0 0 0 0 0 0 0 0 0 0 755 0
Ins 0 0 5 0 0 0 0 0 0 0 0 14 533
=====
    
```

Tabla 6 reconocedor entrenado con personas nativas del idioma Inglés, adaptado con muestras de un Hispano y probado con el mismo Hispano .

Actualmente se han desarrollado reconocedores utilizando frases cortas sin adaptar y adaptados a Hispanos, utilizando las siguientes frases en inglés, pronunciadas por nativos del idioma Inglés y por hispanos:

- a) Yesterday I drove to my brother's house and had lunch with him and his family
- b) Cowboys ride bulls and horses
- c) Baby girls play house but baby boys play with trains
- d) The couple ran on the Sandy beach looking for shells

La gramática utilizada para el reconocedor 1 y 2 fue la siguiente:

```
$palabras = sil | yesterday | i | drove | to | my | brothers  
| house | and | had | lunch | with | he | and | his | family  
| in | miami | it | is | usually | warm | and | never | snows  
| have | to | work | today | but | tomorrow | my | day | off  
| they | watched | the | ship | named | constitution | as |  
she | shipped | below | waves | kansas | states | new |  
basketball | tickets | cost | ten | dollars | now | cowboys  
| ride | big | bulls | horses | couple | ran | on | looking |  
sandy | beach | for | shells | cabbage | patch | kids | are |  
dolls | that | adopt | little | girls | their | mothers |  
baby | play | house | but | boys | trains | we | relaxed |  
passed | through | tiny | mexican | shops;  
(<$palabras>)
```

A continuación se muestran los resultados obtenidos:

- 1) Reconocedor de palabras con modelos nativos del idioma Inglés.

En el primer reconocedor no se realizó ninguna adaptación, es decir se utilizaron Modelos Ocultos de Markov (HMM) entrenados únicamente con hablantes nativos del idioma Inglés (Personas de Estados Unidos) y probado con cuatro personas hispanas. De los resultados obtenidos se tiene un reconocimiento de hispanos del 39.02 %. La matriz de confusión se muestra en el apéndice B.

- 2) Reconocedor de palabras con modelos adaptados a hispanos hablando Inglés.

En el segundo reconocedor se adaptaron los modelos nativos del experimento anterior con muestras de un hablante hispano, pronunciando las frases en inglés. Una vez adaptados los modelos se realizó el reconocimiento con cuatro hablantes hispanos diferentes (incluyendo al hablante con el que se efectuó la

adaptación), y se nota una mejora en el reconocimiento. En esta ocasión se obtuvo un reconocimiento de 53.76%. Así mismo se nota un mayor reconocimiento en el hablante con el que se realizó la adaptación. La gramática utilizada fue la misma del reconocedor (1).

Una vez terminados estos experimentos se efectuaron otros dos reconocedores, en los cuales se modificó la gramática del reconocimiento, reduciendo así la perplejidad del reconocedor, para enfocar la gramática al reconocimiento de frases (en lugar de hacerlo a nivel palabra). La gramática utilizada en estos reconocedores fue la siguiente:

```
$palabra1 = yesterday | cowboys | the | baby;  
$palabra2 = i |ride |couple | girls;  
$palabra3 = drove | big | ran | play;  
$palabra4 = to | bulls | on | house;  
$palabra5 = my | and | the | but;  
$palabra6 = brothers | horses | sandy | baby;  
$palabra7 = house | beach | boys | and | looking | play |  
had | for | with | lunch | shells | trains | with | he | his  
| family |shell;  
([sil] $palabra1 [sil] $palabra2 [sil] $palabra3 [sil]  
$palabra4 [sil] $palabra5 [sil] $palabra6 [sil] <$palabra7  
[sil]>)
```

3) Reconocedor de Frases no adaptado.

En el primer reconocedor se realizó el reconocimiento con los modelos sin adaptar, es decir se utilizaron los modelos entrenados únicamente con hablantes nativos del inglés. En los resultados se observa una mejora en este reconocimiento respecto al que se tuvo con una gramática a nivel palabra. El reconocimiento fue de 51.73%.

4) Reconocedor de Frases Adaptado.

En este reconocedor se realizó una adaptación con muestras de un solo hablante pronunciando las frases en inglés y posteriormente se realizó el reconocimiento de cuatro hablantes (incluyendo al hablante con el que se adaptaron los modelos). De los resultados se puede observar una mejora respecto a los reconocedores a nivel palabra y del reconocedor no adaptado a nivel frase. El porcentaje de reconocimiento fue de 60.69%.

Se puede observar que en los modelos adaptados, el modelo nativo se sensibiliza con un número reducido de muestras, lo cual es suficiente para lograr la adaptación y estos modelos sean capaces de reconocer a nuevos locutores (hispanos).

También es importante notar la mejora obtenida con la gramática, en la que se establece una secuencia de palabras respecto a las frases predefinidas, esto ayuda a que el reconocedor tenga que distinguir en un universo menor de palabras y pueda seleccionar de manera más precisa cuál fue la palabra pronunciada, ya que se debe recordar que básicamente es un proceso probabilístico.

Finalmente se desarrollaron nuevos reconocedores de voz de frases, en donde se agregaron más frases hasta tener un total de 8. En estos reconocedores se varió el hablante utilizado para adaptación, así como el número de hablantes hispanos utilizados para prueba de reconocimiento. Algunos reconocedores se obtuvieron sin realizar adaptación, para poder observar que se tienen mejores resultados de reconocimiento a hispanos cuando se tiene un modelo adaptado a hispanos. Se utilizaron las siguientes frases en inglés, pronunciadas por hispanos:

- a) Yesterday I drove to my brother's house and had lunch with him and his family
- b) I have to work today but tomorrow is my day off
- c) They watched the ship named the Constitution as she slipped below the waves
- d) Kansas state's new basketball tickets cost ten dollars now
- e) Cowboys ride bulls and horses
- f) Baby girls play house but baby boys play with trains
- g) Cabbage Patch kids are adults that adopt little girls as their mothers
- h) The couple ran on the Sandy beach looking for shells

Para este tipo de reconocedores básicamente el reconocimiento se realizó en base a la gramática orientada a frases.

5) Reconocedor de ocho frases Adaptado con un hablante hispano (s2063).

En este reconocedor se realizó una adaptación con muestras de un solo hablante pronunciando las frases en inglés y posteriormente se realizó el reconocimiento de cinco hablantes hispanos (incluyendo al hablante con el que se adaptaron los modelos). El porcentaje de reconocimiento fue de 40.04%. Podemos observar que cuando se realiza un reconocimiento con un número mayor de hablantes hispanos, el porcentaje de reconocimiento tiende a disminuir. Analizando las diferentes muestras de hablantes se nota que esto tiene mucho que ver con la pronunciación del hablante

utilizado en la adaptación, así como la pronunciación de los hablantes utilizados para el reconocimiento.

6) Reconocedor de ocho frases no adaptado.

En este reconocedor se realizó el reconocimiento de los cinco hablantes hispanos pero sin realizar adaptación. De los resultados se puede observar que el porcentaje de reconocimiento fue de 33.78%, lo cual representa un porcentaje menor respecto al modelo adaptado.

7) Reconocedor de ocho frases adaptado con una hablante hispano (s2063) y probado con 24 hablantes hispanos.

En este reconocedor adaptado, se utilizaron los archivos con las ocho frases indicadas anteriormente, pero esta vez se utilizaron 24 archivos para el reconocimiento, incluyendo el archivo con el que se realizó la adaptación. Para este experimento se obtuvo un porcentaje de reconocimiento de 43.98%. Nuevamente podemos observar que mientras mayor es el número de hablantes con los que se realiza la adaptación, el porcentaje de reconocimiento tiende a disminuir, debido a que la pronunciación es muy variada entre los diferentes hispanos.

8) Reconocedor de ocho frases no adaptado y probado con 24 hablantes.

En este reconocedor se utilizaron los archivos con las ocho frases indicadas anteriormente, pero esta vez se utilizarán 24 hablantes hispanos para el reconocimiento, sin adaptación. Para este experimento se obtuvo un porcentaje de reconocimiento de 30.67%

9) Reconocedor de ocho frases adaptado con un hablante hispano (s2015) y probado con 24 hablantes hispanos.

En este reconocedor se utilizaron los archivos con las ocho frases indicadas anteriormente, con 24 hablantes hispanos para el reconocimiento, incluyendo el archivo con el que se realizó la adaptación. A diferencia del reconocedor 7 se utilizó otro hablante diferente para la adaptación (s2015). Para este experimento se obtuvo un porcentaje de reconocimiento de 37%. Se puede observar que dependiendo de la muestra que se utilice para la adaptación, se puede tener un reconocimiento mayor o por el contrario puede disminuir el porcentaje de reconocimiento, esto se observa que va ligado con la pronunciación del hispano seleccionado para adaptar.

10) Reconocedor de ocho frases utilizando dos archivos para adaptación (s2063 y s2053)

En este reconocedor se utilizaron dos hablantes hispanos para la adaptación de los modelos y se probó con los 24 hablantes hispanos, incluyendo los dos hablantes con los que se adaptaron los modelos. El porcentaje de reconocimiento fue de 42.39%. Aún cuando se duplicó el número de hispanos para adaptar los modelos, el resultado es muy similar a l del reconocedor siete, lo cual vuelve a dar indicios de que la muestra o muestras utilizadas en la adaptación, son clave para los resultados del reconocimiento.

En la tabla 7 se muestra el Resumen de resultados de los diferentes experimentos y en el apéndice B se muestran las matrices de confusión de cada reconocedor de voz.

Tipo de Reconocedor	Porcentaje de Reconocimiento
1) Reconocedor de palabras con modelos nativos del idioma Inglés pronunciando cuatro frases cortas y probado con hispanos sin adaptar .	39.02 %.
2) Reconocedor de palabras con modelos adaptados a hispanos hablando Inglés (pronunciando cuatro frases cortas).	53.76%.
3) Reconocedor de cuatro Frases cortas no adaptado (con gramática orientada a frases).	51.73%.
4) Reconocedor de Frases cuatro cortas adaptado (con gramática orientada a frases).	60.69%.
5) Reconocedor de ocho frases Adaptado con un hablante hispano (s2063).	40.04%
6) Reconocedor de ocho frases no adaptado	33.78%.
7) Reconocedor de ocho frases adaptado con una hablante hispano (s2063) y probado con 24 hablantes hispanos.	43.98%
8) Reconocedor de ocho frases no adaptado y probado con 24 hablantes.	30.67%
9) Reconocedor de ocho frases adaptado con un hablante hispano (s2015) y probado con 24 hablantes hispanos.	37.79%
10) Reconocedor de ocho frases utilizando dos archivos para adaptación (s2063 y s2053)	42.39%

Tabla 7. Resumen de resultados de los reconocedores de frases

Capítulo IV: Conclusiones y Trabajo Futuro

4.1 Conclusiones

Existe una gran diversidad de circunstancias que se reflejan en la señal de voz como son: el estado de ánimo, el tono de voz, la pronunciación de diferentes regiones aún en un mismo idioma, el sexo de la persona, entre otras. Todo esto influye en un reconocedor de voz para poder determinar de manera más precisa la palabra pronunciada, el tipo de muestras que se tiene para entrenar y probar si los modelos son representativos para reconocer mejor a un grupo de personas. Si a todo esto se le agrega el realizar un Reconocedor de Voz para personas que pronunciarán en otro idioma y que además son aprendices de dicho idioma, esto hace un reto mayor para el Reconocedor de Voz.

En los diferentes archivos con los que se realizó el desarrollo de los Reconocedores se tienen hablantes hispanos pronunciando mejor el inglés y otros cuya pronunciación es mala, esto también repercute en los resultados del Reconocedor, sin embargo este reconocedor será parte de un Instructor Automatizado, en el que se debe conjuntar el resultado obtenido para poder retroalimentar a los usuarios cuando no están pronunciando adecuadamente.

Otro factor que influye en los resultados del reconocimiento, es la gramática, ya que puede dar la flexibilidad para reconocer palabras aisladas o bien se puede restringir a un grupo de palabras ordenadas en una oración, lo cual reduce el número de alternativas y mejora el porcentaje de reconocimiento, sin embargo, si el hablante no produce la secuencia de palabras como se estableció previamente, provocaba que los resultados del reconocedor fueran más bajos. Esto se observó en algunas de las grabaciones que previamente se tenían para este reconocedor, las personas al darse cuenta que no pronunciaban adecuadamente una palabra, tendían a repetir la palabra sin haber terminado la oración, o bien reiniciaban nuevamente la oración, alterando la secuencia de palabras establecida en la gramática. Esto de acuerdo a la experiencia adquirida se puede mejorar estableciendo una gramática más robusta, en la que permitamos que cada palabra se puede repetir y a su vez demos la opción de repetir la frase. Este experimento se podría realizar en el corto plazo.

El número de muestras para adaptación también puede incrementarse, aunque en este punto va a ser determinante la calidad de la pronunciación de las muestras para lograr un mayor porcentaje de reconocimiento. Dado que este proyecto se

englobará en uno mayor, "Instructor Automatizado", cuando se obtenga un reconocimiento más bajo, esto reflejará una pronunciación cuya calidad sea baja, además de que se pueden analizar la información de resultados de reconocimiento por cada archivo reconocido, en los que se determina cómo fue reconocida cada palabra.

4.2 Trabajo Futuro

Por lo expuesto anteriormente el trabajo futuro se encaminará a utilizar más muestras para la adaptación de modelos (HMM's), hacer mas pruebas de gramática e integrar el sistema al módulo de evaluación de la pronunciación.

En el desarrollo de este trabajo se puede apreciar la gran ayuda que se tiene con la teoría de los Modelos Ocultos de Markov, que en este caso se encuentran aplicados a través de las herramientas del HTK.

Paralelamente con los avances del presente trabajo, se han realizado tres artículos que se han presentado a congresos, y esto ha permitido tener un panorama mayor de las aplicaciones que se puede tener en el Reconocimiento de Voz.

En el área de medicina se podría desarrollar una interfaz de reconocimiento de voz a un sistema de inteligencia artificial para poder proporcionar a un autómata los síntomas de algún padecimiento y este a través de su base de datos orientar el tipo de especialista que puede consultar.

Otras aplicaciones se enfocan al estudio de los formantes de la señal de voz, ya que se considera que son una característica única en cada persona y podría ser una forma de identificación como las huellas digitales.

Encontramos investigadores que están realizando reconocedores para sonidos de delfines.

Como éstos existen un sin número de aplicaciones y estudios de investigación que día con día aumentan y encuentran nuevas aplicaciones en nuestra vida diaria.

BIBLIOGRAFÍA

- [1] Bratt Harry, Neumeyer Leo, Shriberg Elizabeth, and Franco Horacio. Collection and detailed transcription of a speech database for development of language learning technologies. SRI International 1997.
- [2] Eskenazi Maxine. Using a Computer in Foreign Language Pronunciation Training: What Advantages?. VOL 16 No.3. CALICO Journal 1999.
- [3] Franco Horacio. Speech Thechnology and Research Laboratory Language Education". SRI International November 1999.
- [4] Kawi Goh, Hirose Keikichi. Teaching the pronunciation of Japanese double- mora phonemes using speech recognition technology". Speech Communication 30 (2000) 131-143.
- [5] Rabiner Lawrence R., Fellow. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. IEEE 77(2), February 1989,: 257-286.
- [6] Rabiner Lawrence, Biing-Hwang Juang. Fundamentals of Speech Recognition. PTR Prentice Hall 1993.
- [7] M. Antonieta García Galván, René O. Aréchiga Martínez . Adapted speech recognizer to improve English pronunciation of Hispanics. First Pan-American/Iberian Meeting on Acoustics, Cancún, México. 2-6 Diciembre 2002.
- [8] Jonathan E. Hamaker. MLLR: A speaker Adaptation Technique for LVCSR. Department of Electrical and Computer Engineering, Mississippi State University.
- [9] M. A. García Galván, R. Aréchiga Martínez, F.Ramírez Rojas. Reconocedor de Voz. Congreso SOMI XVI, Octubre 2001.
- [10] M. A. García Galván, R. Aréchiga Martínez. Reconocedor de Voz Adaptado para mejorar la pronunciación de Inglés en Hispanos. Congreso SOMI XVIII, Octubre 2003.

APÉNDICE A

Resultados del Reconocedor de Voz Adaptado con frases utilizando dos archivos de hispanos para la adaptación.

```
----- Sentence Scores -----
===== HTK Results Analysis =====
Date: Thu Dec 4 19:39:29 2003
Ref : /home/ggma/corpora/phrasestest_FV2.mlf
Rec : recoutAdapt_FV5.mlf
----- File Results -----
s0023_phr1.rec: 55.88( 50.00) [H= 19, D= 11, S= 4, I= 2, N= 34]
Aligned transcription: s0023_phr1.lab vs s0023_phr1.rec
LAB: yesterday sil i sil drove sil i sil drove sil to sil my sil
brothers sil house sil and sil had sil lunch sil with sil him sil
and sil his sil family sil
REC: sil yesterday sil i drove to the sil
brothers sil with for house sil and had lunch with he and sil
and his family sil
s0023_phr3.rec: 31.82( 27.27) [H= 7, D= 3, S= 12, I= 1, N= 22]
Aligned transcription: s0023_phr3.lab vs s0023_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow
sil is sil my sil day sil off sil
REC: sil the i drove to the horses play sil with sil
for and sil his sil with and house sil
s0023_phr4.rec: 30.77( 23.08) [H= 8, D= 2, S= 16, I= 2, N= 26]
Aligned transcription: s0023_phr4.lab vs s0023_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil
constitution sil and sil she sil shipped sil below sil the sil waves
sil
REC: sil baby ride sil big sil to sil my baby boys his lunch
had sil trains his he with sil with with had sil with
sil
s0023_phr5.rec: 50.00( 5.00) [H= 10, D= 1, S= 9, I= 9, N= 20]
Aligned transcription: s0023_phr5.lab vs s0023_phr5.rec
LAB: kansas sil states sil new sil baseball sil basketball
sil tickets sil cost sil ten sil dollars sil now
sil
REC: sil the i sil play on the sil baby sil he with
sil beach sil for boys boys play beach sil boys sil shell boys sil
house sil
s0023_phr6.rec: 58.33( 41.67) [H= 7, D= 4, S= 1, I= 2, N= 12]
Aligned transcription: s0023_phr6.lab vs s0023_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride big bulls and horses sil he sil
s0023_phr7.rec: 60.00( 55.00) [H= 12, D= 8, S= 0, I= 1, N= 20]
Aligned transcription: s0023_phr7.lab vs s0023_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran on the sandy beach sil
looking for shells sil
s0023_phr8.rec: 32.14( 25.00) [H= 9, D= 3, S= 16, I= 2, N= 28]
Aligned transcription: s0023_phr8.lab vs s0023_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil
adopt sil little sil little sil girls sil and sil as sil
their sil mothers sil
REC: sil the i big to sil but sil baby sil for boys sil
had had looking house sil looking house trains sil house sil
beach house sil
s0023_phr9.rec: 70.00( 65.00) [H= 14, D= 6, S= 0, I= 1, N= 20]
```

Aligned transcription: s0023_phr9.lab vs s0023_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil play house sil but sil baby boys
play with trains sil
s0025_phr1.rec: 43.33(40.00) [H= 13, D= 14, S= 3, I= 1, N= 30]
Aligned transcription: s0025_phr1.lab vs s0025_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil family sil
REC: sil yesterday i drove to my brothers house
sil had had lunch had had and family sil
s0025_phr3.rec: 22.73(18.18) [H= 5, D= 11, S= 6, I= 1, N= 22]
Aligned transcription: s0025_phr3.lab vs s0025_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: sil the i play to sil but
brothers and and sil
s0025_phr4.rec: 19.23(19.23) [H= 5, D= 6, S= 15, I= 0, N= 26]
Aligned transcription: s0025_phr4.lab vs s0025_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil
constitution sil and sil she sil shipped sil below sil the sil waves
sil
REC: sil yesterday sil couple ran on but horses
play shell and sil and shells had sil he house with
his
s0025_phr5.rec: 16.67(16.67) [H= 3, D= 5, S= 10, I= 0, N= 18]
Aligned transcription: s0025_phr5.lab vs s0025_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost sil
ten sil dollars sil now sil
REC: sil the couple sil play to but
brothers house and lunch had sil
s0025_phr6.rec: 75.00(58.33) [H= 9, D= 2, S= 1, I= 2, N= 12]
Aligned transcription: s0025_phr6.lab vs s0025_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride sil big sil house and sil horses he sil
s0025_phr7.rec: 55.00(50.00) [H= 11, D= 8, S= 1, I= 1, N= 20]
Aligned transcription: s0025_phr7.lab vs s0025_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran on the sandy beach sil
looking sil shell sil
s0025_phr8.rec: 8.33(8.33) [H= 2, D= 7, S= 15, I= 0, N= 24]
Aligned transcription: s0025_phr8.lab vs s0025_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt sil
little sil girls sil as sil their sil mothers sil
REC: sil the i ran to and
brothers house had looking and house house and and and sil
s0025_phr9.rec: 50.00(45.00) [H= 10, D= 9, S= 1, I= 1, N= 20]
Aligned transcription: s0025_phr9.lab vs s0025_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls play house but baby boys
sil shell trains sil
s2001_phr1.rec: 62.50(59.38) [H= 20, D= 9, S= 3, I= 1, N= 32]
Aligned transcription: s2001_phr1.lab vs s2001_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil i sil had sil lunch sil with sil him sil and sil his sil
family sil
REC: sil yesterday sil i drove sil to my brothers sil house
sil and had lunch sil with he with and his sil had
sil
s2001_phr3.rec: 22.73(18.18) [H= 5, D= 10, S= 7, I= 1, N= 22]
Aligned transcription: s2001_phr3.lab vs s2001_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: the i ran to but
baby sil with house his lunch trains sil
s2001_phr4.rec: 30.77(23.08) [H= 8, D= 0, S= 18, I= 2, N= 26]

Aligned transcription: s2001_phr4.lab vs s2001_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil constitution sil as sil she sil shipped sil below sil the sil waves sil
REC: baby i ran to the baby shells sil had sil he trains he beach shell sil house sil he sil shell beach sil he house he with sil
s2001_phr5.rec: 27.78(11.11) [H= 5, D= 1, S= 12, I= 3, N= 18]
Aligned transcription: s2001_phr5.lab vs s2001_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost sil ten sil dollars sil now sil
REC: sil the i drove to and sandy and sil house his looking beach sil house sil and lunch house sil
s2001_phr6.rec: 75.00(58.33) [H= 9, D= 1, S= 2, I= 2, N= 12]
Aligned transcription: s2001_phr6.lab vs s2001_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys sil ride sil big sil house and horses sil he sil
s2001_phr8.rec: 29.17(25.00) [H= 7, D= 2, S= 15, I= 1, N= 24]
Aligned transcription: s2001_phr8.lab vs s2001_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt sil little sil girls sil as sil their sil mothers sil
REC: sil the ride sil ran to the sandy house sil had sil and house sil play had shells sil and and house sil
s2001_phr9.rec: 65.00(60.00) [H= 13, D= 6, S= 1, I= 1, N= 20]
Aligned transcription: s2001_phr9.lab vs s2001_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil play sil with sil trains sil
REC: sil baby girls sil play house sil the baby boys sil play with trains sil
s2015_phr1.rec: 83.33(73.33) [H= 25, D= 2, S= 3, I= 3, N= 30]
Aligned transcription: s2015_phr1.lab vs s2015_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house sil and sil had sil lunch sil with sil him sil and sil his sil family sil
REC: sil yesterday sil i drove sil to sil my brothers had lunch sil and sil had sil lunch sil with sil he and sil and sil his sil family beach sil
s2015_phr3.rec: 36.36(31.82) [H= 8, D= 2, S= 12, I= 1, N= 22]
Aligned transcription: s2015_phr3.lab vs s2015_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is sil my sil day sil off sil
REC: sil the i drove sil on sil the baby sil lunch sil play sil lunch his play beach sil for sil
s2015_phr4.rec: 46.15(15.38) [H= 12, D= 3, S= 11, I= 8, N= 26]
Aligned transcription: s2015_phr4.lab vs s2015_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil constitution sil as sil she sil shipped sil below sil the sil waves sil
REC: sil baby i ran house sil the sil baby shells sil had and sil beach looking sil for shell sil had sil his he sil beach shell sil had with sil
s2015_phr5.rec: 44.44(-5.56) [H= 8, D= 1, S= 9, I= 9, N= 18]
Aligned transcription: s2015_phr5.lab vs s2015_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost sil ten sil dollars sil now sil
REC: sil the i ran house and sil baby sil he and with house with looking sil beach sil house sil and lunch sil trains had sil
s2015_phr6.rec: 100.00(83.33) [H= 12, D= 0, S= 0, I= 2, N= 12]
Aligned transcription: s2015_phr6.lab vs s2015_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys sil ride sil big sil bulls sil and sil horses his sil
s2015_phr7.rec: 63.64(59.09) [H= 14, D= 5, S= 3, I= 1, N= 22]
Aligned transcription: s2015_phr7.lab vs s2015_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil looking sil for sil the sil shells sil

REC: sil the sil couple sil ran sil on the sandy beach sil
looking for beach shell his sil
s2015_phr8.rec: 37.50(20.83) [H= 9, D= 1, S= 14, I= 4, N= 24]
Aligned transcription: s2015_phr8.lab vs s2015_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil
adopt sil little sil girls sil as sil their sil mothers sil
REC: sil cowboys sil couple sil big sil on sil the horses sil
had and had boys beach sil he he had had sil he trains lunch sil
s2015_phr9.rec: 75.00(65.00) [H= 15, D= 3, S= 2, I= 2, N= 20]
Aligned transcription: s2015_phr9.lab vs s2015_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil play house sil but sil baby boys sil
trains sil with sil play and sil
s2016_phr1.rec: 61.76(50.00) [H= 21, D= 9, S= 4, I= 4, N= 34]
Aligned transcription: s2016_phr1.lab vs s2016_phr1.rec
LAB: i sil yesterday sil drove sil to sil my sil brother sil my sil
brothers sil house sil and sil had sil lunch sil with sil
he sil and sil his sil family sil
REC: sil yesterday sil i drove sil to my
brothers sil trains house sil and had sil play house sil play sil his sil
he sil and he family sil
s2016_phr3.rec: 31.82(31.82) [H= 7, D= 3, S= 12, I= 0, N= 22]
Aligned transcription: s2016_phr3.lab vs s2016_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: sil the ride sil play sil on the baby sil boys sil
for lunch sil his play trains sil
s2016_phr4.rec: 46.15(-11.54) [H= 12, D= 3, S= 11, I= 15, N= 26]
Aligned transcription: s2016_phr4.lab vs s2016_phr4.rec
LAB: they sil watched sil the sil ship sil
named sil the sil constitution sil as sil she
sil shipped sil below sil the sil waves sil
REC: sil the i ran on and sandy sil beach he sil he and
and sil beach boys he boys shell sil had sil had sil house beach
sil he trains sil with for sil beach and with sil
s2016_phr5.rec: 50.00(11.11) [H= 9, D= 0, S= 9, I= 7, N= 18]
Aligned transcription: s2016_phr5.lab vs s2016_phr5.rec
LAB: kansas sil states sil new sil basketball sil
tickets sil cost sil ten sil dollars sil now sil

REC: sil yesterday i sil big sil bulls sil but sil baby had
boys sil he trains sil house sil play and lunch sil house sil
s2016_phr6.rec: 83.33(41.67) [H= 10, D= 0, S= 2, I= 5, N= 12]
Aligned transcription: s2016_phr6.lab vs s2016_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil
horses sil
REC: sil cowboys sil ride sil big sil to and sil horses sil and sil house
he sil
s2016_phr8.rec: 50.00(20.83) [H= 12, D= 1, S= 11, I= 7, N= 24]
Aligned transcription: s2016_phr8.lab vs s2016_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil
that sil adopt sil little sil girls sil as sil their sil
mothers sil
REC: sil baby i ran house sil the horses sil he sil house sil
boys sil and sil his sil for sil with and sil and sil his boys and
sil
s2016_phr9.rec: 72.73(68.18) [H= 16, D= 5, S= 1, I= 1, N= 22]
Aligned transcription: s2016_phr9.lab vs s2016_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil train sil trains sil
REC: sil baby girls sil play house sil but sil baby boys sil
play with sil play trains sil
s2029_phr1.rec: 50.00(40.00) [H= 15, D= 12, S= 3, I= 3, N= 30]
Aligned transcription: s2029_phr1.lab vs s2029_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil
family sil

REC: sil yesterday i drove to my brothers house
 sil had and his had had sil and his sil had and
 sil
 s2029_phr3.rec: 13.64(13.64) [H= 3, D= 8, S= 11, I= 0, N= 22]
 Aligned transcription: s2029_phr3.lab vs s2029_phr3.rec
 LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
 sil my sil day sil off sil
 REC: sil the i play on the sil baby
 sil had house had play house
 s2029_phr4.rec: 18.75(18.75) [H= 6, D= 7, S= 19, I= 0, N= 32]
 Aligned transcription: s2029_phr4.lab vs s2029_phr4.rec
 LAB: they sil watched sil the sil ship sil named sil the sil constitution
 sil as sil she sil shipped sil as sil she sil shipped sil below
 sil the sil waves sil
 REC: the i ran on the horses
 shell play with house sil he shells his his his sil his his his
 sil had play his sil
 s2029_phr5.rec: 11.11(11.11) [H= 2, D= 5, S= 11, I= 0, N= 18]
 Aligned transcription: s2029_phr5.lab vs s2029_phr5.rec
 LAB: kansas sil states sil new sil basketball sil tickets sil cost sil
 ten sil dollars sil now sil
 REC: the girls play sil to but baby
 his house play house had sil
 s2029_phr6.rec: 50.00(50.00) [H= 6, D= 3, S= 3, I= 0, N= 12]
 Aligned transcription: s2029_phr6.lab vs s2029_phr6.rec
 LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
 REC: cowboys ride sil big sil house the horses he
 s2029_phr7.rec: 45.00(40.00) [H= 9, D= 8, S= 3, I= 1, N= 20]
 Aligned transcription: s2029_phr7.lab vs s2029_phr7.rec
 LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
 looking sil for sil shells sil
 REC: sil the couple ran on and sandy beach had
 his for shells sil
 s2029_phr8.rec: 4.17(4.17) [H= 1, D= 8, S= 15, I= 0, N= 24]
 Aligned transcription: s2029_phr8.lab vs s2029_phr8.rec
 LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt
 sil little sil girls sil as sil their sil mothers sil
 REC: cowboys i play to and horses
 with and sil he and house his shell had his
 s2029_phr9.rec: 55.00(50.00) [H= 11, D= 7, S= 2, I= 1, N= 20]
 Aligned transcription: s2029_phr9.lab vs s2029_phr9.rec
 LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
 play sil with sil trains sil
 REC: sil baby girls play house sil the baby boys sil
 he with sil trains
 s2031_phr1.rec: 50.00(46.43) [H= 14, D= 8, S= 6, I= 1, N= 28]
 Aligned transcription: s2031_phr1.lab vs s2031_phr1.rec
 LAB: yesterday sil i sil drove sil my sil brothers sil house sil and
 sil had sil lunch sil with sil him sil and sil his sil family sil
 REC: sil yesterday sil i drove to my brothers house
 sil had sil house sil with shell sil house for with sil
 s2031_phr3.rec: 22.73(18.18) [H= 5, D= 12, S= 5, I= 1, N= 22]
 Aligned transcription: s2031_phr3.lab vs s2031_phr3.rec
 LAB: i sil have sil to sil work sil today sil but sil tomorrow sil
 is sil my sil day sil off sil
 REC: sil the couple play to sil but
 horses play trains sil
 s2031_phr4.rec: 34.62(26.92) [H= 9, D= 4, S= 13, I= 2, N= 26]
 Aligned transcription: s2031_phr4.lab vs s2031_phr4.rec
 LAB: they sil watched sil the sil ship sil named sil the sil
 constitution sil as sil she sil shipped sil below sil the sil
 waves sil
 REC: sil the i ran house sil the sandy and
 sil his his he shell sil lunch his had sil he house sil with sil
 s2031_phr5.rec: 38.89(5.56) [H= 7, D= 1, S= 10, I= 6, N= 18]
 Aligned transcription: s2031_phr5.lab vs s2031_phr5.rec
 LAB: kansas sil states sil new sil basketball sil tickets
 sil cost sil ten sil dollars sil now sil

REC: sil the i ran to the sandy sil he and sil boys
sil boys sil he trains sil house and for house sil
s2031_phr6.rec: 41.67(25.00) [H= 5, D= 1, S= 6, I= 2, N= 12]
Aligned transcription: s2031_phr6.lab vs s2031_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys sil i ran to but sil horses sil had house sil
s2031_phr8.rec: 37.50(29.17) [H= 9, D= 2, S= 13, I= 2, N= 24]
Aligned transcription: s2031_phr8.lab vs s2031_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil
adopt sil little sil girls sil as sil their sil mothers sil
REC: sil cowboys sil ride sil big to sil but horses sil for and
house sil play had sil house sil house his house sil
s2031_phr9.rec: 55.00(50.00) [H= 11, D= 6, S= 3, I= 1, N= 20]
Aligned transcription: s2031_phr9.lab vs s2031_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby sil girls sil ran house sil the baby boys
sil trains trains sil
s2049_phr1.rec: 63.33(60.00) [H= 19, D= 6, S= 5, I= 1, N= 30]
Aligned transcription: s2049_phr1.lab vs s2049_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil family sil
REC: sil yesterday sil i drove sil to my brothers house
sil had had sil he house with sil he sil and his sil had sil
s2049_phr3.rec: 40.91(31.82) [H= 9, D= 2, S= 11, I= 2, N= 22]
Aligned transcription: s2049_phr3.lab vs s2049_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil
is sil my sil day sil off sil
REC: sil baby sil ride sil play sil on sil the baby sil with sil
for lunch sil his play play for sil
s2049_phr4.rec: 43.33(16.67) [H= 13, D= 1, S= 16, I= 8, N= 30]
Aligned transcription: s2049_phr4.lab vs s2049_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the
sil constitution sil as sil she sil shipped sil below sil
below sil the sil the sil waves sil
REC: sil baby ride sil big sil house and sil baby sil and house
sil he play shell and sil house sil shell shell he he sil
beach sil beach house sil with lunch sil shell with sil
s2049_phr5.rec: 50.00(11.11) [H= 9, D= 1, S= 8, I= 7, N= 18]
Aligned transcription: s2049_phr5.lab vs s2049_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets
sil cost sil ten sil dollars sil now sil
REC: sil the i sil drove house sil and sil baby lunch
sil with for sil his beach sil house sil and house his sil
s2049_phr6.rec: 66.67(58.33) [H= 8, D= 1, S= 3, I= 1, N= 12]
Aligned transcription: s2049_phr6.lab vs s2049_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride sil big sil bulls sil the sandy house sil
s2049_phr7.rec: 65.00(55.00) [H= 13, D= 5, S= 2, I= 2, N= 20]
Aligned transcription: s2049_phr7.lab vs s2049_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil baby sil ride sil ran on the sandy beach sil
had looking sil for shells sil
s2049_phr8.rec: 37.50(25.00) [H= 9, D= 0, S= 15, I= 3, N= 24]
Aligned transcription: s2049_phr8.lab vs s2049_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil
adopt sil little sil girls sil as sil their sil mothers sil
REC: cowboys sil ride sil big sil on sil the horses sil had sil had
lunch sil he with and had sil house had trains he lunch sil
s2049_phr9.rec: 75.00(70.00) [H= 15, D= 4, S= 1, I= 1, N= 20]
Aligned transcription: s2049_phr9.lab vs s2049_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil play house sil but sil baby boys sil
play with sil play sil
s2050_phr1.rec: 47.22(44.44) [H= 17, D= 11, S= 8, I= 1, N= 36]
Aligned transcription: s2050_phr1.lab vs s2050_phr1.rec

LAB: yesterday sil i sil drove sil my sil brothers sil house sil to
sil my sil brothers sil house sil and sil had sil lunch sil with sil him
sil and sil his sil family sil

REC: sil yesterday sil i big on my brothers house
sil play and and house sil and had sil his and with
he and sil family sil
s2050_phr3.rec: 22.73(18.18) [H= 5, D= 8, S= 9, I= 1, N= 22]

Aligned transcription: s2050_phr3.lab vs s2050_phr3.rec

LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil

REC: sil the i ran to but sil
baby and house his had play house sil
s2050_phr4.rec: 30.77(19.23) [H= 8, D= 6, S= 12, I= 3, N= 26]

Aligned transcription: s2050_phr4.lab vs s2050_phr4.rec

LAB: they sil watched sil the sil ship sil named sil the
sil constitution sil as sil she sil shipped sil below sil the sil
waves sil

REC: sil baby ride sil play house the baby sil beach
house sil he shell sil his his beach beach sil
house play sil

s2050_phr5.rec: 50.00(27.78) [H= 9, D= 1, S= 8, I= 4, N= 18]

Aligned transcription: s2050_phr5.lab vs s2050_phr5.rec

LAB: kansas sil states sil new sil basketball sil tickets sil cost
sil ten sil dollars sil now sil

REC: sil the girls sil big sil to but sil horses sil he
sil his sil house sil and house house sil

s2050_phr6.rec: 41.67(8.33) [H= 5, D= 1, S= 6, I= 4, N= 12]

Aligned transcription: s2050_phr6.lab vs s2050_phr6.rec

LAB: cowboys sil ride sil big sil bulls sil and sil horses
sil

REC: sil the i ran sil on sil my baby for for sil house his
sil

s2050_phr7.rec: 55.00(50.00) [H= 11, D= 6, S= 3, I= 1, N= 20]

Aligned transcription: s2050_phr7.lab vs s2050_phr7.rec

LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil

REC: sil the couple ran sil on the sandy beach sil had
he for shell sil

s2050_phr8.rec: 38.46(30.77) [H= 10, D= 1, S= 15, I= 2, N= 26]

Aligned transcription: s2050_phr8.lab vs s2050_phr8.rec

LAB: cabbage sil patch sil kids sil are sil are sil dolls sil that
sil adopt sil little sil girls sil as sil their sil mothers sil

REC: sil the ride sil ran sil to the sandy sil and sil boys
sil had sil had and beach and had sil house sil had had sil

s2050_phr9.rec: 70.00(65.00) [H= 14, D= 4, S= 2, I= 1, N= 20]

Aligned transcription: s2050_phr9.lab vs s2050_phr9.rec

LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil

REC: sil baby ride sil play house sil but sil baby boys sil
play with sil play sil

s2053_phr1.rec: 58.82(58.82) [H= 20, D= 12, S= 2, I= 0, N= 34]

Aligned transcription: s2053_phr1.lab vs s2053_phr1.rec

LAB: yesterday sil i sil drove sil to sil my sil brothers sil house sil
and sil had sil lunch sil with sil him sil and sil and sil his sil his sil
family sil

REC: yesterday i drove sil to my brothers house sil
and had lunch with he sil had and his sil his
family sil

s2053_phr3.rec: 9.09(9.09) [H= 2, D= 8, S= 12, I= 0, N= 22]

Aligned transcription: s2053_phr3.lab vs s2053_phr3.rec

LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil

REC: sil the i drove bulls the baby
play for his play beach house sil

s2053_phr4.rec: 34.62(7.69) [H= 9, D= 1, S= 16, I= 7, N= 26]

Aligned transcription: s2053_phr4.lab vs s2053_phr4.rec

LAB: they sil watched sil the sil ship sil named sil the
 sil constitution sil as sil she sil shipped sil below
 sil the sil waves sil
 REC: sil the i ran on the sandy sil beach he play sil play
 with sil had and house he play shell sil house he sil his he
 sil beach had with sil
 s2053_phr5.rec: 22.22(16.67) [H= 4, D= 1, S= 13, I= 1, N= 18]
 Aligned transcription: s2053_phr5.lab vs s2053_phr5.rec
 LAB: kansas sil states sil new sil basketball sil tickets sil cost
 sil ten sil dollars sil now sil
 REC: sil the i ran house and sandy boys with he his
 sil house sil and house house sil
 s2053_phr6.rec: 91.67(75.00) [H= 11, D= 1, S= 0, I= 2, N= 12]
 Aligned transcription: s2053_phr6.lab vs s2053_phr6.rec
 LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
 REC: sil cowboys sil ride sil big sil bulls sil and horses he sil
 s2053_phr7.rec: 70.00(65.00) [H= 14, D= 6, S= 0, I= 1, N= 20]
 Aligned transcription: s2053_phr7.lab vs s2053_phr7.rec
 LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
 looking sil for sil shells sil
 REC: sil the couple ran sil on the sandy sil beach sil
 looking for shells sil
 s2053_phr8.rec: 29.17(12.50) [H= 7, D= 1, S= 16, I= 4, N= 24]
 Aligned transcription: s2053_phr8.lab vs s2053_phr8.rec
 LAB: cabbage sil patch sil kids sil are sil dolls sil that
 sil adopt sil little sil girls sil and sil their sil mothers sil
 REC: the i big sil to and sil baby his sil had boys with and
 sil had looking with and house sil had and and house sil
 s2053_phr9.rec: 70.83(66.67) [H= 17, D= 6, S= 1, I= 1, N= 24]
 Aligned transcription: s2053_phr9.lab vs s2053_phr9.rec
 LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
 play sil with sil train sil with sil trains sil
 REC: sil baby girls sil play house sil but baby boys sil
 play with trains sil with sil trains sil
 s2055_phr1.rec: 53.33(50.00) [H= 16, D= 11, S= 3, I= 1, N= 30]
 Aligned transcription: s2055_phr1.lab vs s2055_phr1.rec
 LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
 sil and sil had sil lunch sil with sil him sil and sil his sil family sil
 REC: sil yesterday sil i drove to but sil brothers house
 sil and had lunch with he lunch sil and sil
 s2055_phr3.rec: 9.09(9.09) [H= 2, D= 15, S= 5, I= 0, N= 22]
 Aligned transcription: s2055_phr3.lab vs s2055_phr3.rec
 LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
 sil my sil day sil off sil
 REC: the couple play to but
 baby house
 s2055_phr4.rec: 38.46(30.77) [H= 10, D= 5, S= 11, I= 2, N= 26]
 Aligned transcription: s2055_phr4.lab vs s2055_phr4.rec
 LAB: they sil watched sil the sil ship sil named sil the sil
 constitution sil as sil she sil shipped sil below sil the sil
 waves sil
 REC: sil the ride sil big sil to the baby sil and
 sil beach with and sil his sil with he looking with sil
 s2055_phr5.rec: 16.67(16.67) [H= 3, D= 6, S= 9, I= 0, N= 18]
 Aligned transcription: s2055_phr5.lab vs s2055_phr5.rec
 LAB: kansas sil states sil new sil basketball sil tickets sil cost sil ten
 sil dollars sil now sil
 REC: sil the ride sil play to but
 horses and house his sil
 s2055_phr6.rec: 58.33(50.00) [H= 7, D= 3, S= 2, I= 1, N= 12]
 Aligned transcription: s2055_phr6.lab vs s2055_phr6.rec
 LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
 REC: cowboys ride sil big sil on the horses he sil
 s2055_phr7.rec: 55.00(50.00) [H= 11, D= 8, S= 1, I= 1, N= 20]
 Aligned transcription: s2055_phr7.lab vs s2055_phr7.rec
 LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
 looking sil for sil shells sil

REC: sil the couple ran on the sandy beach
looking his sil shells sil
s2055_phr8.rec: 16.67(16.67) [H= 4, D= 12, S= 8, I= 0, N= 24]
Aligned transcription: s2055_phr8.lab vs s2055_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt sil
little sil girls sil as sil their sil mothers sil
REC: sil the couple drove sil to sil but brothers trains had
sil
s2055_phr9.rec: 65.00(60.00) [H= 13, D= 7, S= 0, I= 1, N= 20]
Aligned transcription: s2055_phr9.lab vs s2055_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls play house sil but baby boys
play with sil trains sil
s2059_phr1.rec: 64.29(53.57) [H= 18, D= 5, S= 5, I= 3, N= 28]
Aligned transcription: s2059_phr1.lab vs s2059_phr1.rec
LAB: yesterday sil i sil drove sil my sil brothers sil house sil
and sil had sil lunch sil with sil he sil and sil his sil family
sil
REC: sil yesterday couple ran sil on sil my brothers house sil
play had lunch sil with he and sil and sil his sil for lunch sil
s2059_phr3.rec: 27.27(27.27) [H= 6, D= 5, S= 11, I= 0, N= 22]
Aligned transcription: s2059_phr3.lab vs s2059_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil
is sil my sil day sil off sil
REC: sil the couple play sil bulls sil the baby lunch sil his he
trains sil for sil
s2059_phr4.rec: 42.31(-7.69) [H= 11, D= 3, S= 12, I= 13, N= 26]
Aligned transcription: s2059_phr4.lab vs s2059_phr4.rec
LAB: they sil watched sil the sil ship sil
named sil the sil constitution sil as sil she sil shipped
sil below sil the sil waves sil
REC: sil the i ran on my sandy sil beach he play he sil with and house
sil looking shell and sil house sil he sil shell he beach he
sil he lunch sil he with sil
s2059_phr5.rec: 38.89(-5.56) [H= 7, D= 1, S= 10, I= 8, N= 18]
Aligned transcription: s2059_phr5.lab vs s2059_phr5.rec
LAB: kansas sil states sil new sil basketball sil
tickets sil cost sil ten sil dollars sil now sil
REC: sil the i ran house sil the sil baby sil he and house with
looking his sil house had sil play and lunch house sil
s2059_phr6.rec: 41.67(25.00) [H= 5, D= 1, S= 6, I= 2, N= 12]
Aligned transcription: s2059_phr6.lab vs s2059_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys sil i ran to the baby lunch sil and house sil
s2059_phr7.rec: 65.00(60.00) [H= 13, D= 5, S= 2, I= 1, N= 20]
Aligned transcription: s2059_phr7.lab vs s2059_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple sil ran sil on the sandy beach sil
looking for shell and sil
s2059_phr8.rec: 29.17(12.50) [H= 7, D= 1, S= 16, I= 4, N= 24]
Aligned transcription: s2059_phr8.lab vs s2059_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil
that sil adopt sil little sil girls sil as sil their sil mothers sil
REC: sil the girls big sil house the sil baby his sil and
lunch he had looking he he and and sil house sil beach and lunch
sil
s2059_phr9.rec: 75.00(65.00) [H= 15, D= 3, S= 2, I= 2, N= 20]
Aligned transcription: s2059_phr9.lab vs s2059_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys
sil play sil with sil trains sil
REC: sil baby i sil play house sil but sil baby boys he
sil trains sil with sil trains sil
s2060_phr1.rec: 63.33(60.00) [H= 19, D= 7, S= 4, I= 1, N= 30]
Aligned transcription: s2060_phr1.lab vs s2060_phr1.rec

LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil family
sil
REC: sil yesterday sil i drove to my brothers sil house
sil play had lunch sil with trains sil had sil his sil and
sil
s2060_phr3.rec: 22.73(13.64) [H= 5, D= 8, S= 9, I= 2, N= 22]
Aligned transcription: s2060_phr3.lab vs s2060_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil
is sil my sil day sil off sil
REC: sil the i ran sil to but
baby play for and his play play house sil
s2060_phr4.rec: 30.77(30.77) [H= 8, D= 3, S= 15, I= 0, N= 26]
Aligned transcription: s2060_phr4.lab vs s2060_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil
constitution sil as sil she sil shipped sil below sil the sil
waves sil
REC: sil baby ride sil big sil to and sil baby
sil his beach and sil lunch beach trains sil beach boys with sil
s2060_phr5.rec: 33.33(33.33) [H= 6, D= 1, S= 11, I= 0, N= 18]
Aligned transcription: s2060_phr5.lab vs s2060_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost
sil ten sil dollars sil now sil
REC: sil yesterday sil i sil drove to the horses his
sil house sil and house had sil
s2060_phr6.rec: 83.33(66.67) [H= 10, D= 1, S= 1, I= 2, N= 12]
Aligned transcription: s2060_phr6.lab vs s2060_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride sil big sil bulls sil the sil horses he sil
s2060_phr7.rec: 55.00(50.00) [H= 11, D= 7, S= 2, I= 1, N= 20]
Aligned transcription: s2060_phr7.lab vs s2060_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran on the sandy beach sil had
he for shells sil
s2060_phr8.rec: 29.17(29.17) [H= 7, D= 3, S= 14, I= 0, N= 24]
Aligned transcription: s2060_phr8.lab vs s2060_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt
sil little sil girls sil as sil their sil mothers sil
REC: sil cowboys sil ride sil big sil to and horses
had and he had sil house sil his family his sil
s2060_phr9.rec: 75.00(70.00) [H= 15, D= 5, S= 0, I= 1, N= 20]
Aligned transcription: s2060_phr9.lab vs s2060_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil play house sil but sil baby boys sil
play with trains sil
s2063_phr1.rec: 42.50(40.00) [H= 17, D= 16, S= 7, I= 1, N= 40]
Aligned transcription: s2063_phr1.lab vs s2063_phr1.rec
LAB: yesterday sil i sil drove sil my sil brothers sil house sil i sil
drove sil to sil my sil brothers sil house sil and sil had sil lunch
sil with sil him sil and sil his sil family sil
REC: sil yesterday i drove to my brothers house
sil play and house lunch house house sil and had lunch
with and and his family sil
s2063_phr3.rec: 27.27(18.18) [H= 6, D= 8, S= 8, I= 2, N= 22]
Aligned transcription: s2063_phr3.lab vs s2063_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil
is sil my sil day sil off sil
REC: sil the sil i ran to but baby and
house sil his play play house sil
s2063_phr4.rec: 25.00(12.50) [H= 6, D= 3, S= 15, I= 3, N= 24]
Aligned transcription: s2063_phr4.lab vs s2063_phr4.rec
LAB: they sil watched sil the sil ship sil named sil
constitution sil as sil she sil shipped sil below sil the sil
waves sil

REC: sil the i ran house sil the sil baby shell play and house looking shell and sil had shells had beach looking trains sil
 s2063_phr4.rec: 25.00(12.50) [H= 6, D= 3, S= 15, I= 3, N= 24]
 Aligned transcription: s2063_phr4.lab vs s2063_phr4.rec
 LAB: they sil watched sil the sil ship sil named sil constitution sil as sil she sil shipped sil below sil the sil waves sil

REC: sil the i ran house sil the sil baby shell play and house looking shell and sil had shells had beach looking trains sil
 s2063_phr6.rec: 50.00(41.67) [H= 6, D= 4, S= 2, I= 1, N= 12]
 Aligned transcription: s2063_phr6.lab vs s2063_phr6.rec
 LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
 REC: sil cowboys ride big sil bulls the horses he
 s2063_phr7.rec: 59.09(54.55) [H= 13, D= 9, S= 0, I= 1, N= 22]
 Aligned transcription: s2063_phr7.lab vs s2063_phr7.rec
 LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil looking sil for sil shell sil shells sil
 REC: sil the couple ran on the sandy beach sil looking for shell sil shells
 s2063_phr8.rec: 16.67(16.67) [H= 4, D= 4, S= 16, I= 0, N= 24]
 Aligned transcription: s2063_phr8.lab vs s2063_phr8.rec
 LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt sil little sil girls sil as sil their sil mothers sil
 REC: sil the i big to and sandy sil and house and and had had house sil and and lunch sil
 s2063_phr9.rec: 65.00(60.00) [H= 13, D= 7, S= 0, I= 1, N= 20]
 Aligned transcription: s2063_phr9.lab vs s2063_phr9.rec
 LAB: baby sil girls sil play sil house sil but sil baby sil boys sil play sil with sil trains sil
 REC: sil baby girls play house sil but sil baby boys play with sil trains
 s2064_phr1.rec: 46.67(43.33) [H= 14, D= 14, S= 2, I= 1, N= 30]
 Aligned transcription: s2064_phr1.lab vs s2064_phr1.rec
 LAB: yesterday sil i sil drove sil to sil my sil brothers sil house sil and sil had sil lunch sil with sil him sil and sil his sil family sil
 REC: sil yesterday i drove sil to my brothers house sil and lunch his and sil had sil
 s2064_phr3.rec: 9.09(9.09) [H= 2, D= 10, S= 10, I= 0, N= 22]
 Aligned transcription: s2064_phr3.lab vs s2064_phr3.rec
 LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is sil my sil day sil off sil
 REC: sil the i play on the baby play boys with and sil
 s2064_phr4.rec: 23.08(19.23) [H= 6, D= 6, S= 14, I= 1, N= 26]
 Aligned transcription: s2064_phr4.lab vs s2064_phr4.rec
 LAB: they sil watched sil the sil ship sil named sil the sil constitution sil as sil she sil shipped sil below sil the sil waves sil
 REC: the girls big sil to the baby with house beach shell sil and shells he with sil beach for with sil
 s2064_phr5.rec: 16.67(16.67) [H= 3, D= 2, S= 13, I= 0, N= 18]
 Aligned transcription: s2064_phr5.lab vs s2064_phr5.rec
 LAB: kansas sil states sil new sil basketball sil tickets sil cost sil ten sil dollars sil now sil
 REC: the girls play to the brothers with with his sil house sil and lunch house sil
 s2064_phr6.rec: 41.67(33.33) [H= 5, D= 5, S= 2, I= 1, N= 12]
 Aligned transcription: s2064_phr6.lab vs s2064_phr6.rec
 LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
 REC: cowboys i big bulls the horses his sil
 s2064_phr7.rec: 55.00(50.00) [H= 11, D= 8, S= 1, I= 1, N= 20]
 Aligned transcription: s2064_phr7.lab vs s2064_phr7.rec
 LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil looking sil for sil shells sil
 REC: sil the couple sil ran on the sandy beach looking had shells sil
 s2064_phr8.rec: 16.67(16.67) [H= 4, D= 8, S= 12, I= 0, N= 24]
 Aligned transcription: s2064_phr8.lab vs s2064_phr8.rec

LAB: cabbage sil patch sil kids sil are sil dolls sil that sil
adopt sil little sil girls sil as sil their sil mothers sil
REC: sil cowboys couple drove
bulls the brothers sil he with house his had sil boys sil
s2064_phr9.rec: 55.00(50.00) [H= 11, D= 6, S= 3, I= 1, N= 20]
Aligned transcription: s2064_phr9.lab vs s2064_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil play house sil the baby boys
sil trains sil play sil
s2083_phr1.rec: 53.33(50.00) [H= 16, D= 10, S= 4, I= 1, N= 30]
Aligned transcription: s2083_phr1.lab vs s2083_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil family
sil
REC: sil yesterday i drove sil to my brothers house
sil and had lunch sil with he looking his and and
sil
s2083_phr3.rec: 13.64(13.64) [H= 3, D= 13, S= 6, I= 0, N= 22]
Aligned transcription: s2083_phr3.lab vs s2083_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: the couple play to but
horses had house sil
s2083_phr4.rec: 36.67(20.00) [H= 11, D= 5, S= 14, I= 5, N= 30]
Aligned transcription: s2083_phr4.lab vs s2083_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil
constitution sil and sil she sil shipped sil she
sil shipped sil below sil the sil waves sil
REC: sil the i ran house sil the sandy sil
play and sil beach house beach shell and sil his his sil beach
his his sil his and with sil
s2083_phr5.rec: 22.22(22.22) [H= 4, D= 3, S= 11, I= 0, N= 18]
Aligned transcription: s2083_phr5.lab vs s2083_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost sil
ten sil dollars sil now sil
REC: sil the girls sil play to but baby beach sil
house and lunch house sil
s2083_phr6.rec: 75.00(58.33) [H= 9, D= 2, S= 1, I= 2, N= 12]
Aligned transcription: s2083_phr6.lab vs s2083_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys sil ride sil big sil bulls the horses he sil
s2083_phr7.rec: 50.00(45.00) [H= 10, D= 7, S= 3, I= 1, N= 20]
Aligned transcription: s2083_phr7.lab vs s2083_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran on the sandy beach had his
sil his shells sil
s2083_phr8.rec: 21.43(21.43) [H= 6, D= 3, S= 19, I= 0, N= 28]
Aligned transcription: s2083_phr8.lab vs s2083_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt sil
little sil girls sil as sil their sil as sil their sil mothers sil
REC: sil the ride sil ran to and sandy boys sil had had
house sil he he and house and beach sil with and house sil
s2083_phr9.rec: 65.00(60.00) [H= 13, D= 6, S= 1, I= 1, N= 20]
Aligned transcription: s2083_phr9.lab vs s2083_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil play house sil the baby boys sil
play with trains sil
s2084_phr1.rec: 53.33(46.67) [H= 16, D= 11, S= 3, I= 2, N= 30]
Aligned transcription: s2084_phr1.lab vs s2084_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil he sil and sil his sil
family sil
REC: sil yesterday sil i drove to my brothers sil house
and had lunch play his sil he trains for and
sil

s2084_phr3.rec: 13.64(9.09) [H= 3, D= 10, S= 9, I= 1, N= 22]
Aligned transcription: s2084_phr3.lab vs s2084_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: the i ran to but
baby play for his he and play house
s2084_phr4.rec: 30.00(23.33) [H= 9, D= 4, S= 17, I= 2, N= 30]
Aligned transcription: s2084_phr4.lab vs s2084_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil
constitution sil as sil she sil as sil she sil shipped sil below sil
the sil waves sil
REC: baby ride sil big sil to the sil baby and with his
he shell trains he sil and sil he sil beach sil he and with
with sil
s2084_phr5.rec: 22.22(22.22) [H= 4, D= 2, S= 12, I= 0, N= 18]
Aligned transcription: s2084_phr5.lab vs s2084_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost sil
ten sil dollars sil now sil
REC: sil the girls sil big sil to but baby
looking his his and looking and sil
s2084_phr6.rec: 66.67(50.00) [H= 8, D= 1, S= 3, I= 2, N= 12]
Aligned transcription: s2084_phr6.lab vs s2084_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride sil big sil on sil but horses sil he sil
s2084_phr7.rec: 55.00(50.00) [H= 11, D= 6, S= 3, I= 1, N= 20]
Aligned transcription: s2084_phr7.lab vs s2084_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple sil ran on and sil sandy beach
play beach for shells sil
s2084_phr8.rec: 17.86(17.86) [H= 5, D= 6, S= 17, I= 0, N= 28]
Aligned transcription: s2084_phr8.lab vs s2084_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt
sil little sil girls sil as sil their sil as sil their sil
mothers sil
REC: sil cowboys couple sil big sil on
and horses and and sil play looking looking and play looking and and
lunch sil
s2084_phr9.rec: 70.00(65.00) [H= 14, D= 5, S= 1, I= 1, N= 20]
Aligned transcription: s2084_phr9.lab vs s2084_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby sil girls sil play house sil the baby boys sil
play with trains sil
s2085_phr1.rec: 46.67(43.33) [H= 14, D= 12, S= 4, I= 1, N= 30]
Aligned transcription: s2085_phr1.lab vs s2085_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil family sil
REC: sil yesterday i sil drove to my brothers house
sil and had had with he his his had sil
s2085_phr3.rec: 18.18(18.18) [H= 4, D= 10, S= 8, I= 0, N= 22]
Aligned transcription: s2085_phr3.lab vs s2085_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: the couple play to sil but
baby lunch his play trains sil
s2085_phr4.rec: 23.08(23.08) [H= 6, D= 5, S= 15, I= 0, N= 26]
Aligned transcription: s2085_phr4.lab vs s2085_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil constitution
sil as sil she sil shipped sil below sil the sil waves sil
REC: sil baby ride sil big to the sil
baby boys he play shell house his he sil he shell with sil
s2085_phr5.rec: 27.78(11.11) [H= 5, D= 1, S= 12, I= 3, N= 18]
Aligned transcription: s2085_phr5.lab vs s2085_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost
sil ten sil dollars sil now sil
REC: sil the i big house and sandy with his with he
sil his sil house play sil house house sil

s2085_phr6.rec: 66.67(50.00) [H= 8, D= 1, S= 3, I= 2, N= 12]
Aligned transcription: s2085_phr6.lab vs s2085_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride sil big sil bulls sil the baby boys his sil
s2085_phr7.rec: 50.00(45.00) [H= 10, D= 9, S= 1, I= 1, N= 20]
Aligned transcription: s2085_phr7.lab vs s2085_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran on the sandy beach
family for shells sil
s2085_phr8.rec: 33.33(25.00) [H= 8, D= 1, S= 15, I= 2, N= 24]
Aligned transcription: s2085_phr8.lab vs s2085_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that
sil adopt sil little sil girls sil as sil their sil mothers sil
REC: sil the ride sil ran to the sil horses had had looking
he had sil shells sil and sil play and had sil his sil
s2085_phr9.rec: 65.00(60.00) [H= 13, D= 6, S= 1, I= 1, N= 20]
Aligned transcription: s2085_phr9.lab vs s2085_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil big house but sil baby boys sil
play with trains sil
s2089_phr1.rec: 46.88(40.62) [H= 15, D= 14, S= 3, I= 2, N= 32]
Aligned transcription: s2089_phr1.lab vs s2089_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil i sil had sil lunch sil with sil him sil and sil his sil
family sil
REC: sil yesterday i drove to my brothers house
sil and and sil lunch had play lunch his sil
and sil
s2089_phr3.rec: 27.27(22.73) [H= 6, D= 9, S= 7, I= 1, N= 22]
Aligned transcription: s2089_phr3.lab vs s2089_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: the i ran sil to but
baby and house sil his play trains sil
s2089_phr4.rec: 30.77(11.54) [H= 8, D= 4, S= 14, I= 5, N= 26]
Aligned transcription: s2089_phr4.lab vs s2089_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the
sil constitution sil and sil she sil shipped sil below sil the sil
waves sil
REC: sil the i ran on but baby shells sil play and house
he beach shell and sil had shells sil with sil he for
play his sil
s2089_phr5.rec: 16.67(16.67) [H= 3, D= 3, S= 12, I= 0, N= 18]
Aligned transcription: s2089_phr5.lab vs s2089_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost
sil ten sil dollars sil now sil
REC: sil the girls sil play to but horses
had his house and house house sil
s2089_phr6.rec: 50.00(50.00) [H= 6, D= 2, S= 4, I= 0, N= 12]
Aligned transcription: s2089_phr6.lab vs s2089_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: cowboys ride sil big sil on the sandy house sil
s2089_phr7.rec: 55.00(50.00) [H= 11, D= 9, S= 0, I= 1, N= 20]
Aligned transcription: s2089_phr7.lab vs s2089_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran on the sandy beach
looking for shells sil
s2089_phr8.rec: 25.00(25.00) [H= 7, D= 4, S= 17, I= 0, N= 28]
Aligned transcription: s2089_phr8.lab vs s2089_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt
sil little sil girls sil as sil their sil as sil their sil mothers sil
REC: cowboys sil ride sil big to and sil horses had
sil house sil beach and house his his sil had and and and sil
s2089_phr9.rec: 75.00(70.00) [H= 15, D= 5, S= 0, I= 1, N= 20]
Aligned transcription: s2089_phr9.lab vs s2089_phr9.rec

LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls play house sil but sil baby boys sil
play with sil trains sil
s2099_phr1.rec: 56.67(53.33) [H= 17, D= 9, S= 4, I= 1, N= 30]
Aligned transcription: s2099_phr1.lab vs s2099_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil family sil
REC: sil yesterday i drove to my brothers house
sil he sil had sil lunch with sil he and and had family sil
s2099_phr3.rec: 18.18(18.18) [H= 4, D= 8, S= 10, I= 0, N= 22]
Aligned transcription: s2099_phr3.lab vs s2099_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is sil
my sil day sil off sil
REC: sil the i ran sil bulls sil the
baby play house boys trains sil
s2099_phr4.rec: 30.77(26.92) [H= 8, D= 5, S= 13, I= 1, N= 26]
Aligned transcription: s2099_phr4.lab vs s2099_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil
constitution sil as sil she sil shipped sil below sil the sil waves
sil
REC: the i drove sil to and sil baby sil house
beach he sil house his sil with sil he house play sil
s2099_phr5.rec: 27.78(27.78) [H= 5, D= 2, S= 11, I= 0, N= 18]
Aligned transcription: s2099_phr5.lab vs s2099_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost sil
ten sil dollars sil now sil
REC: sil the girls play to but horses his house sil
had sil boys sil house sil
s2099_phr6.rec: 66.67(50.00) [H= 8, D= 2, S= 2, I= 2, N= 12]
Aligned transcription: s2099_phr6.lab vs s2099_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride sil big sil on the sil horses he sil
s2099_phr7.rec: 55.00(50.00) [H= 11, D= 7, S= 2, I= 1, N= 20]
Aligned transcription: s2099_phr7.lab vs s2099_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran on the sandy beach sil
with his for shells sil
s2099_phr8.rec: 20.83(20.83) [H= 5, D= 4, S= 15, I= 0, N= 24]
Aligned transcription: s2099_phr8.lab vs s2099_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt
sil little sil girls sil as sil their sil mothers sil
REC: sil the couple drove to and sandy sil house had
sil house sil play and house with and lunch sil
s2099_phr9.rec: 60.00(55.00) [H= 12, D= 7, S= 1, I= 1, N= 20]
Aligned transcription: s2099_phr9.lab vs s2099_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil play house but baby boys sil
play with play sil
s2112_phr1.rec: 40.00(36.67) [H= 12, D= 16, S= 2, I= 1, N= 30]
Aligned transcription: s2112_phr1.lab vs s2112_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil family sil
REC: sil yesterday i drove to my brothers house
had and with and his and sil
s2112_phr3.rec: 13.64(13.64) [H= 3, D= 8, S= 11, I= 0, N= 22]
Aligned transcription: s2112_phr3.lab vs s2112_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: sil the i drove sil bulls the
baby play house with he house sil
s2112_phr4.rec: 23.08(23.08) [H= 6, D= 5, S= 15, I= 0, N= 26]
Aligned transcription: s2112_phr4.lab vs s2112_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil constitution
sil as sil she sil shipped sil below sil the sil waves sil

REC: sil the sil ride big to and sil baby
sil with his play shells beach he shell play play his sil
s2112_phr5.rec: 16.67(16.67) [H= 3, D= 4, S= 11, I= 0, N= 18]
Aligned transcription: s2112_phr5.lab vs s2112_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost sil
ten sil dollars sil now sil
REC: sil the ride play to but baby beach sil
house and house had sil
s2112_phr6.rec: 66.67(50.00) [H= 8, D= 3, S= 1, I= 2, N= 12]
Aligned transcription: s2112_phr6.lab vs s2112_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride sil big sil bulls the horses he sil
s2112_phr7.rec: 45.00(40.00) [H= 9, D= 8, S= 3, I= 1, N= 20]
Aligned transcription: s2112_phr7.lab vs s2112_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran on the sandy beach
he beach he shells sil
s2112_phr8.rec: 8.33(8.33) [H= 2, D= 6, S= 16, I= 0, N= 24]
Aligned transcription: s2112_phr8.lab vs s2112_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt sil
little sil girls sil as sil their sil mothers sil
REC: sil the i play to the baby
house had and play and house house shell house and sil
s2112_phr9.rec: 60.00(55.00) [H= 12, D= 7, S= 1, I= 1, N= 20]
Aligned transcription: s2112_phr9.lab vs s2112_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls play house sil the baby boys
play sil with trains sil
s2128_phr1.rec: 63.33(60.00) [H= 19, D= 9, S= 2, I= 1, N= 30]
Aligned transcription: s2128_phr1.lab vs s2128_phr1.rec
LAB: yesterday sil i sil drove sil to sil my sil brothers sil house
sil and sil had sil lunch sil with sil him sil and sil his sil family sil
REC: sil yesterday i drove sil to my brothers house
sil and had sil lunch with he and sil and his family sil
s2128_phr3.rec: 22.73(18.18) [H= 5, D= 10, S= 7, I= 1, N= 22]
Aligned transcription: s2128_phr3.lab vs s2128_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: the i big sil to but
baby looking boys had play house sil
s2128_phr4.rec: 15.38(15.38) [H= 4, D= 4, S= 18, I= 0, N= 26]
Aligned transcription: s2128_phr4.lab vs s2128_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil
constitution sil as sil she sil shipped sil below sil the sil waves
sil
REC: the i ran to my horses shell play and
sil his house and shell beach his had sil for play his sil
s2128_phr5.rec: 16.67(11.11) [H= 3, D= 2, S= 13, I= 1, N= 18]
Aligned transcription: s2128_phr5.lab vs s2128_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost
sil ten sil dollars sil now sil
REC: sil the i ran to and sandy house boys had his
sil house and house had sil
s2128_phr6.rec: 83.33(75.00) [H= 10, D= 1, S= 1, I= 1, N= 12]
Aligned transcription: s2128_phr6.lab vs s2128_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: cowboys sil ride sil big sil on sil and horses he sil
s2128_phr7.rec: 65.00(60.00) [H= 13, D= 6, S= 1, I= 1, N= 20]
Aligned transcription: s2128_phr7.lab vs s2128_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple sil ran sil on the sandy beach he
looking for shells sil
s2128_phr8.rec: 25.00(20.83) [H= 6, D= 3, S= 15, I= 1, N= 24]
Aligned transcription: s2128_phr8.lab vs s2128_phr8.rec

LAB: cabbage sil patch sil kids sil are sil dolls sil that sil adopt
sil little sil girls sil as sil their sil mothers sil
REC: sil the i big to and sil baby he sil house sil had
sil had house sil looking house his family his
s2128_phr9.rec: 70.00(65.00) [H= 14, D= 6, S= 0, I= 1, N= 20]
Aligned transcription: s2128_phr9.lab vs s2128_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls sil play house sil but baby boys sil
play with trains sil
s2130_phr1.rec: 38.24(35.29) [H= 13, D= 13, S= 8, I= 1, N= 34]
Aligned transcription: s2130_phr1.lab vs s2130_phr1.rec
LAB: yesterday sil i sil drove sil my sil brothers sil house sil my
sil brothers sil house sil and sil had sil lunch sil with sil him sil and
sil his sil family sil
REC: sil yesterday i drove on the horses house
sil house house sil and sil had lunch with
had play house had sil
s2130_phr3.rec: 22.73(18.18) [H= 5, D= 10, S= 7, I= 1, N= 22]
Aligned transcription: s2130_phr3.lab vs s2130_phr3.rec
LAB: i sil have sil to sil work sil today sil but sil tomorrow sil is
sil my sil day sil off sil
REC: sil the i ran to but
baby play house sil play had sil
s2130_phr4.rec: 30.77(23.08) [H= 8, D= 5, S= 13, I= 2, N= 26]
Aligned transcription: s2130_phr4.lab vs s2130_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the sil
constitution sil as sil she sil shipped sil below sil the sil
waves sil
REC: sil the i ran on the sandy shell
sil play and sil house sil had his shell had house sil his play sil
s2130_phr5.rec: 27.78(27.78) [H= 5, D= 3, S= 10, I= 0, N= 18]
Aligned transcription: s2130_phr5.lab vs s2130_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil cost sil
ten sil dollars sil now sil
REC: sil the i play sil to but sil horses
house sil and house house sil
s2130_phr6.rec: 66.67(50.00) [H= 8, D= 3, S= 1, I= 2, N= 12]
Aligned transcription: s2130_phr6.lab vs s2130_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: sil cowboys ride sil big sil on and horses he sil
s2130_phr7.rec: 50.00(45.00) [H= 10, D= 7, S= 3, I= 1, N= 20]
Aligned transcription: s2130_phr7.lab vs s2130_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran sil on the sandy beach
sil with his house sil
s2130_phr8.rec: 33.33(29.17) [H= 8, D= 3, S= 13, I= 1, N= 24]
Aligned transcription: s2130_phr8.lab vs s2130_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls sil that sil
adopt sil little sil girls sil as sil their sil mothers sil
REC: sil the i ran sil house the sandy sil and house sil
had house sil looking house sil and sil lunch sil
s2130_phr9.rec: 65.00(60.00) [H= 13, D= 6, S= 1, I= 1, N= 20]
Aligned transcription: s2130_phr9.lab vs s2130_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil
play sil with sil trains sil
REC: sil baby girls play house sil the baby boys sil
play with sil trains sil
s1018_phr4.rec: 30.77(15.38) [H= 8, D= 2, S= 16, I= 4, N= 26]
Aligned transcription: s1018_phr4.lab vs s1018_phr4.rec
LAB: they sil watched sil the sil ship sil named sil the
sil constitution sil as sil she sil shipped sil below sil the sil
waves sil
REC: sil baby i drove sil bulls and sil baby and house he
play shell sil house sil he his he with sil he house sil
play and sil
s1018_phr5.rec: 40.00(5.00) [H= 8, D= 1, S= 11, I= 7, N= 20]

Aligned transcription: s1018_phr5.lab vs s1018_phr5.rec
LAB: kansas sil states sil new sil basketball sil tickets sil
cost sil ten sil dollars sil dollars sil now sil
REC: sil the i big to and sil baby and and with sil
he trains sil house sil and house sil boys lunch sil had and sil
s1018_phr6.rec: 75.00(66.67) [H= 9, D= 2, S= 1, I= 1, N= 12]
Aligned transcription: s1018_phr6.lab vs s1018_phr6.rec
LAB: cowboys sil ride sil big sil bulls sil and sil horses sil
REC: cowboys sil ride sil big bulls and horses sil he sil
s1018_phr7.rec: 60.00(45.00) [H= 12, D= 5, S= 3, I= 3, N= 20]
Aligned transcription: s1018_phr7.lab vs s1018_phr7.rec
LAB: the sil couple sil ran sil on sil the sil sandy sil beach sil
looking sil for sil shells sil
REC: sil the couple ran sil on but sandy beach sil
looking for sil shell sil for house sil
s1018_phr8.rec: 37.50(12.50) [H= 9, D= 1, S= 14, I= 6, N= 24]
Aligned transcription: s1018_phr8.lab vs s1018_phr8.rec
LAB: cabbage sil patch sil kids sil are sil dolls
sil that sil adopt sil little sil girls sil as sil their sil
mothers sil
REC: sil cowboys sil girls sil play to and horses sil and and
sil and and sil play he had play had sil house sil play and and
trains sil
s1018_phr9.rec: 75.00(75.00) [H= 15, D= 4, S= 1, I= 0, N= 20]
Aligned transcription: s1018_phr9.lab vs s1018_phr9.rec
LAB: baby sil girls sil play sil house sil but sil baby sil boys sil play
sil with sil trains sil
REC: baby girls sil play house sil but sil baby sil boys play
looking sil trains sil

```

----- Overall Results -----
SENT: %Correct=0.00 [H=0, S=187, N=187]
WORD: %Corr=42.39, Acc=34.25 [H=1739, D=988, S=1375, I=334, N=4102]
-----
Confusion Matrix
-----
a b b b b b b c c d f f g h h h h h i l l m o p r r s s s s t t t w y
n a e i o r u u o o r a o i a e i o o o u y n l a i a h h i h o r i e
d b a g y o l t u w o m r r r d s r u o n a n d n e e l e a t t s
y c s t l p b v i l s s s k c y e d l l l e i h t i h t
h h s l o e l s e e e i h y l l l n n e Del [%c / %e]
adop 3 1 0 0 0 0 1 0 0 0 0 0 0 0 8 0 0 2 1 0 1 1 0 1 1 0 0 0 0 0 0 0 0 3 [0.0/0.5]
and 47 1 0 0 0 0 0 0 0 0 0 0 0 1 0 1 2 0 1 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 16 [77.0/0.3]
are 1 0 0 2 0 0 0 0 1 0 0 0 0 0 0 0 0 1 2 1 1 1 0 0 1 1 1 0 0 0 2 3 0 0 0 6 [0.0/0.5]
as 6 0 2 0 1 0 0 0 0 0 0 0 0 0 0 0 4 3 4 0 14 0 2 0 0 0 0 4 0 0 0 0 0 0 4 [0.0/1.1]
baby 0 48 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
base 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 [0.0/0.0]
bask 4 1 0 1 0 0 0 0 2 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 3 0 0 0 1 1 0 1 0 3 [0.0/0.5]
beac 0 0 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
belo 0 0 3 0 0 0 0 0 0 0 0 0 0 0 1 0 3 5 2 0 1 0 0 0 0 0 0 0 0 2 2 0 0 0 1 3 0 3 [0.0/0.6]
big 0 0 0 21 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 [87.5/0.1]
boys 0 0 0 0 24 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
brot 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
brot 1 0 0 0 0 0 22 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 [88.0/0.1]
bull 0 1 0 0 0 0 0 11 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 1 0 2 0 0 0 0 0 0 0 0 0 7 [64.7/0.1]
but 0 2 0 0 0 0 0 1 29 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 1 0 0 0 0 0 0 1 0 9 [76.3/0.2]
cabb 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 22 [0.0/0.0]
cons 7 2 1 0 0 0 0 0 0 0 0 0 0 0 2 1 1 1 2 0 0 0 0 2 0 0 0 0 3 0 0 0 0 0 0 0 3 [0.0/0.5]
cost 0 1 5 0 1 0 0 1 0 0 0 0 0 1 0 0 2 5 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 2 0 0 1 [0.0/0.5]
coup 0 0 0 0 0 0 0 0 0 20 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 [95.2/0.0]
cowb 0 0 0 0 0 0 0 0 0 0 23 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
day 1 0 1 0 1 0 1 0 0 0 0 0 0 0 0 2 1 4 2 2 0 0 0 0 0 4 0 0 0 0 0 0 0 1 2 0 2 [0.0/0.5]
doll 14 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 0 0 0 3 0 0 2 0 0 0 0 0 0 0 0 0 [0.0/0.6]
doll 5 1 0 1 1 0 0 0 1 0 0 0 0 0 1 1 0 2 2 2 0 0 0 1 0 0 0 2 0 0 0 1 1 0 0 0 2 [0.0/0.5]
drov 0 0 0 1 0 0 0 0 0 0 21 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 [87.5/0.1]
fami 5 0 0 0 0 0 0 0 0 0 0 9 0 0 0 5 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 2 [42.9/0.3]
for 0 0 1 0 0 0 0 0 0 0 0 0 14 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 3 [77.8/0.1]
girl 3 0 2 0 0 0 0 0 0 0 0 0 0 1 22 5 2 0 0 2 1 2 0 0 0 2 0 0 1 0 0 0 0 1 0 0 4 [50.0/0.5]
had 0 0 0 0 0 0 0 0 0 0 0 0 0 21 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 [95.5/0.0]
have 0 1 0 1 0 0 0 0 0 4 0 0 0 0 0 0 0 0 0 0 0 0 0 4 0 0 0 0 2 0 0 0 0 2 0 0 0 9 [0.0/0.3]
he 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
him 3 0 0 0 0 0 0 0 0 0 0 0 0 0 2 3 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 10 [0.0/0.2]
his 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 15 0 1 0 0 0 0 0 0 0 0 0 6 [83.3/0.1]
hors 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 5 2 13 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 [54.2/0.3]
hous 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 49 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 [98.0/0.0]
i 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 26 0 0 0 0 0 1 0 0 0 0 0 0 0 0 23 [96.3/0.0]
in 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
is 1 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 2 0 1 1 0 0 0 0 0 1 0 0 9 [0.0/0.3]
it 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
kans 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 22 [0.0/0.0]
kids 1 0 0 2 0 0 0 0 0 2 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 4 3 0 0 0 1 1 0 0 0 7 [0.0/0.4]
litt 3 0 1 0 0 1 0 0 0 0 1 0 0 0 4 1 1 1 7 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 2 0 2 [0.0/0.6]
look 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 0 1 0 0 0 12 0 0 0 1 0 0 0 0 0 0 0 0 1 0 3 [66.7/0.1]
lunc 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 3 [80.0/0.1]
mexi 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
    
```


APÉNDICE B

Matrices de Confusión de los Reconocedores de frases

Tabla de Resumen de Resultados de los diferentes Reconocedores utilizando frases cortas.

1) Reconocedor de palabras con modelos nativos del idioma Inglés pronunciando cuatro frases cortas y probado con hispanos sin adaptar .	39.02 %.
2) Reconocedor de palabras con modelos adaptados a hispanos hablando Inglés (pronunciando cuatro frases cortas).	53.76%.
3) Reconocedor de cuatro Frases cortas no adaptado (con gramática orientada a frases).	51.73%.
4) Reconocedor de Frases cuatro cortas adaptado (con gramática orientada a frases).	60.69%.
5) Reconocedor de ocho frases Adaptado con un hablante hispano (s2063).	40.04%
6) Reconocedor de ocho frases no adaptado	33.78%.
7) Reconocedor de ocho frases adaptado con una hablante hispano (s2063) y probado con 24 hablantes hispanos.	43.98%
8) Reconocedor de ocho frases no adaptado y probado con 24 hablantes.	30.67%
9) Reconocedor de ocho frases adaptado con un hablante hispano (s2015) y probado con 24 hablantes hispanos.	37.79%
10) Reconocedor de ocho frases utilizando dos archivos para adaptación (s2063 y s2053)	42.39%

UNIVERSIDAD AUTONOMA METROPOLITANA AZC.
 Maestría en Ciencias de la Computación Reconocedor de Voz Adaptado

===== HTK Results Analysis =====

Date: Fri Sep 19 17:34:06 2003
 Ref : /home/ggma/corpora/phrasetest_sil.mlf
 Rec : recoutAdapt_sinadfrases.mlf

----- Overall Results -----

SENT: %Correct=0.00 [H=0, S=16, N=16]
 WORD: %Corr=51.73, Acc=44.80 [H=179, D=116, S=51, I=24, N=346]

----- Confusion Matrix -----

	a	b	b	b	b	b	c	c	d	f	f	g	h	h	h	h	h	h	i	l	m	o	p	r	r	s	s	s	s	t	t	t	w	y	Del	[%c / %e]		
and	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	[87.5/0.3]	
baby	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2		
beac	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
big	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
boys	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
brot	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[80.0/0.3]	
bull	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	[66.7/0.3]	
but	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
coup	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
cowb	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
drov	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[80.0/0.3]
fami	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[25.0/0.9]
for	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
girl	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
had	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
he	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[75.0/0.3]
his	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	[0.0/0.3]
hors	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
hous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
look	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[0.0/1.2]
lunc	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
my	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	[75.0/0.3]
on	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
play	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	2	
ran	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	
ride	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	
sand	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	
shel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	
shel	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[75.0/0.3]
sil	6	0	0	0	0	0	0	2	0	0	0	0	3	0	5	3	3	0	2	0	1	0	1	0	0	0	0	2	0	47	2	0	1	0	0	95	[60.3/9.0]	
the	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0	0	0	[75.0/0.6]		
to	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
traí	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
with	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	[57.1/0.9]
yest	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Ins	2	0	0	0	0	0	0	0	0	0	0	1	0	2	3	1	0	2	0	0	0	0	0	0	0	0	0	0	11	2	0	0	0	0	0	0		

UNIVERSIDAD AUTONOMA METROPOLITANA AZC.
 Maestría en Ciencias de la Computación Reconocedor de Voz Adaptado

===== HTK Results Analysis =====

Date: Thu Nov 27 20:00:03 2003
 Ref : /home/ggma/corpora/phrasetest_FV1.mlf
 Rec : recoutAdapt_FV1.mlf

----- Overall Results -----
 SENT: %Correct=0.00 [H=0, S=40, N=40]
 WORD: %Corr=40.04, Acc=34.45 [H=358, D=266, S=270, I=50, N=894]

	Confusion Matrix																																								
	a	b	b	b	b	b	b	c	c	d	f	f	g	h	h	h	h	i	l	l	m	o	p	r	r	s	s	s	s	t	t	t	w	y							
	n	a	e	i	o	r	u	u	o	o	r	a	o	i	a	e	h	i	o	o	o	u	y	n	l	a	n	d	n	e	e	l	e	l	e	h	o	r	i	e	
	d	b	a	g	y	o	l	t	u	w	o	m	r	r	d		s	r	u	o	n		a	n	d	n	e	e	l	e	l	e		i	h	t					
	y	c	s	t	l		p	b	v	i	l		s	s	s		e	s	s	k	c		y		e	y	l	l				n	e						Del [%c / %e]		
adop	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2 [0.0/0.3]		
and	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 [85.7/0.2]		
are	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	2 [0.0/0.3]			
as	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1 [0.0/1.1]		
baby	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
base	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [0.0/0.1]		
bask	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 [0.0/0.3]		
beac	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
belo	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	2	0	0	0	0	0	0	0	0 [0.0/0.7]			
big	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
boys	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
brot	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
brot	0	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [83.3/0.1]		
bull	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0		
but	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2 [87.5/0.1]		
cabb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5		
cons	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [0.0/0.7]		
cost	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1 [0.0/0.3]			
coup	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
cowb	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
day	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2 [0.0/0.3]		
doll	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0 [0.0/0.4]			
doll	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1 [0.0/0.4]				
drov	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0		
fami	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2 [66.7/0.1]		
for	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1 [75.0/0.1]			
girl	2	0	0	0	0	0	0	0	0	0	0	1	5	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [50.0/0.6]		
had	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
have	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2 [0.0/0.3]		
he	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
him	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3 [0.0/0.1]		
his	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	
hors	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
hous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0 [90.9/0.1]			
i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5 [71.4/0.2]			
in	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
is	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4 [0.0/0.1]			
it	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
kans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	

UNIVERSIDAD AUTONOMA METROPOLITANA AZC.
 Maestría en Ciencias de la Computación Reconocedor de Voz Adaptado

----- Overall Results -----
 SENT: %Correct=0.00 [H=0, S=187, N=187]
 WORD: %Corr=43.98, Acc=35.76 [H=1804, D=970, S=1328, I=337, N=4102]

	a	b	b	b	b	b	c	c	d	f	f	g	h	h	h	h	h	i	l	l	m	o	p	r	r	s	s	s	s	t	t	t	w	y			
adop	1	0	0	0	0	1	1	0	1	0	0	0	1	0	5	0	1	0	3	0	1	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	5 [0.0/0.5]
and	36	1	0	0	0	0	0	0	0	0	0	0	1	2	0	7	0	0	0	0	0	1	4	0	0	0	1	0	0	4	0	3	0	0	17 [60.0/0.6]		
are	2	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	3	6	0	0	7 [0.0/0.4]			
as	7	0	2	0	0	0	1	0	0	0	0	0	0	0	2	3	6	1	9	0	0	2	0	0	5	0	0	4	1	0	0	0	2	1	3 [0.0/1.1]		
baby	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 [97.9/0.0]		
base	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [0.0/0.0]		
bask	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	3	0	1	3	0	2	0	0	0	2	3	0	1	0	2 [0.0/0.5]			
beac	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
belo	3	0	5	0	0	0	0	0	0	0	0	0	0	3	3	0	1	0	0	0	0	3	0	0	0	2	2	0	0	0	0	3	0	1 [0.0/0.6]			
big	0	1	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0 [87.5/0.1]			
boys	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
brot	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
brot	0	1	0	0	0	20	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	2 [83.3/0.1]				
bull	0	0	0	0	2	0	14	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	2	0	0	0	0	3 [66.7/0.2]				
but	1	3	0	0	1	1	0	25	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	3	0	0	1	0	8 [64.1/0.3]			
cabb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	22 [0.0/0.0]			
cons	3	3	1	0	0	0	0	0	0	0	0	0	1	0	3	0	4	0	0	0	1	2	0	0	0	4	0	0	1	1	0	0	1	1 [0.0/0.6]			
cost	1	0	5	0	0	0	1	0	0	0	0	1	0	9	1	1	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	1 [0.0/0.5]			
coup	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0 [95.2/0.0]			
cowb	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
day	1	0	1	0	2	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	8	0	0	1	0	0	0	0	1	3	0	3 [0.0/0.5]				
doll	7	0	0	0	2	0	0	0	0	0	1	0	0	1	0	0	4	0	0	2	0	0	3	0	0	1	0	0	0	1	1	0	1 [0.0/0.6]				
doll	3	3	0	0	1	0	1	1	0	0	0	0	0	2	0	0	0	1	1	0	0	1	1	0	0	1	0	0	2	0	0	0	6 [0.0/0.4]				
drov	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2			
fami	3	0	1	0	0	0	0	0	0	5	2	0	1	3	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	2	0	3 [25.0/0.4]				
for	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	2 [84.2/0.1]					
girl	3	2	1	0	0	0	1	0	0	0	1	19	3	1	0	0	5	1	1	0	0	1	0	4	0	1	0	0	0	3	0	1	1 [40.4/0.7]				
had	1	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2 [95.2/0.0]					
have	0	1	0	0	0	0	4	1	2	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	1	0	0	0	11 [0.0/0.3]					
he	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
him	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0	13 [0.0/0.2]					
his	0	0	0	0	0	0	0	0	0	0	0	0	1	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	7 [82.4/0.1]					
hors	0	0	0	0	0	0	0	0	0	0	0	0	8	2	10	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [41.7/0.3]			
hous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0 [96.0/0.0]			
i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	21 [89.7/0.1]					
in	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
is	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	1	3	0	0	0	0	0	2	0	0	1	0	12 [0.0/0.3]					
it	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
kans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	21 [0.0/0.0]					
kids	0	0	0	2	0	0	1	2	0	0	0	0	0	0	0	1	3	0	0	0	0	0	4	2	0	0	1	1	0	0	7 [0.0/0.4]						
litt	2	0	2	0	0	1	1	0	0	1	0	3	1	0	1	2	0	0	3	0	0	1	0	0	0	0	1	1	1	0	3 [0.0/0.5]						
look	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	3	2 [78.9/0.1]						
lunc	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	3 [95.0/0.0]							

UNIVERSIDAD AUTONOMA METROPOLITANA AZC.
 Maestría en Ciencias de la Computación Reconocedor de Voz Adaptado

----- Overall Results -----
 SENT: %Correct=0.00 [H=0, S=187, N=187]
 WORD: %Corr=37.79, Acc=30.55 [H=1550, D=1170, S=1382, I=297, N=4102]

		Confusion Matrix																																			
		a	b	b	b	b	b	c	c	d	f	f	g	h	h	h	h	h	i	l	l	m	o	p	r	r	s	s	s	s	t	t	t	w	y		
		n	a	e	i	o	r	u	u	o	o	r	a	o	i	a	e	i	o	o	o	u	n	y	n	l	a	i	a	h	h	i	h	o	r	i	e
		d	b	a	g	y	o	l	t	u	w	o	m	r	r	d	s	r	r	u	o	n	o	y	n	l	a	i	a	h	h	i	h	o	r	i	e
		y	c	h	s	t	l	p	b	v	i	l	s	s	s	s	k	c	c	y	e	d	d	l	l	l	l	l	l	l	l	l	l	l	l	l	l
adop	6	0	0	0	0	1	1	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	2	0	1	1	0	0	
and	44	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	2	0	0	0	1	0	0	0	2	1	5	1	0	
are	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	2	0	0	0	2	4	2	0	0	
as	7	1	0	0	0	0	0	2	0	0	1	0	1	0	1	1	4	0	10	0	1	3	0	0	1	0	0	1	2	0	0	0	5	2	0	0	
baby	0	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
base	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
bask	0	0	0	1	0	3	0	1	0	0	0	0	1	2	0	0	0	1	0	0	0	0	0	2	0	0	2	0	0	0	4	0	1	3	0	0	
beac	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
belo	5	0	1	0	0	0	0	0	0	0	0	1	0	0	3	7	0	2	0	0	1	0	0	0	0	0	0	0	2	0	0	0	2	2	0	0	
big	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
boys	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
brot	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
brot	2	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
bull	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0
but	1	0	0	0	1	0	0	29	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
cabb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
cons	2	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	2	0	0	0	0	1	0	0	0	1	0	2	1	3	0	0	0	5	1	0	
cost	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	1	4	1	4	0	0	2	1	0	0	0	0	0	0	0	0	2	2	0	0	0	
coup	0	0	0	0	0	0	0	0	19	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cowb	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
day	2	0	0	0	1	0	0	0	0	0	0	1	0	4	0	1	1	0	0	1	1	0	0	1	1	0	0	5	0	0	0	1	0	0	0	2	1
doll	13	0	0	0	1	0	0	0	0	0	0	0	3	0	1	0	2	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	
doll	3	1	0	0	0	0	0	0	0	0	0	2	1	0	0	0	1	0	0	0	0	1	0	0	1	1	2	0	0	0	0	0	1	3	0	0	
drov	0	0	0	0	0	0	0	0	0	0	21	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
fami	10	0	0	0	0	0	0	0	0	0	0	6	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
for	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0
girl	3	0	0	0	0	0	0	0	0	0	0	2	24	1	1	1	1	0	3	0	0	2	0	1	2	0	0	0	1	0	1	0	1	0	1	4	
had	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
have	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	
he	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
him	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	
his	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
hors	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	17	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	48	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	19	
in	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
is	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	1	5	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	
it	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
kans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	
kids	0	0	0	0	0	0	1	0	0	0	0	0	5	0	0	0	2	0	0	0	2	0	0	0	1	0	1	0	1	0	0	0	2	0	0	0	
litt	4	0	1	0	0	0	0	0	0	0	1	0	0	1	0	2	0	2	0	5	0	0	2	0	0	1	0	0	0	0	0	1	0	0	0	0	
look	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lunc	2	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4

UNIVERSIDAD AUTONOMA METROPOLITANA AZC.
 Maestría en Ciencias de la Computación Reconocedor de Voz Adaptado

----- Overall Results -----
 SENT: %Correct=0.00 [H=0, S=187, N=187]
 WORD: %Corr=42.39, Acc=34.25 [H=1739, D=988, S=1375, I=334, N=4102]

		Confusion Matrix																																				
		a	b	b	b	b	b	b	c	c	d	f	f	g	h	h	h	h	h	i	l	l	m	o	p	r	r	s	s	s	s	t	t	t	w	y		
		n	a	e	i	o	r	u	u	o	o	r	a	o	i	a	e	i	o	o	o	u	y	n	l	a	i	a	h	h	i	h	o	r	i	e		
		d	b	a	g	y	o	l	t	u	w	o	m	r	r	d	e	s	r	u	o	n				a	n	d	n	e	e	l	e			a	t	s
		y	c	h	s	t	l	p	b	v	i	l	s					e	e	e	i	h			y	e	d	l	l				i	h	t			
																																						Del [%c / %e]
adop	3	1	0	0	0	0	1	0	0	0	0	0	0	0	8	0	0	2	1	0	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	3 [0.0/0.5]		
and	47	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	4	0	1	0	16 [77.0/0.3]	
are	1	0	0	2	0	0	0	1	0	0	0	0	0	0	0	1	2	1	1	1	0	0	0	1	0	1	1	1	0	0	0	2	3	0	0	6 [0.0/0.5]		
as	6	0	2	0	1	0	0	0	0	0	0	0	0	0	0	4	3	4	0	14	0	2	0	0	0	0	0	0	4	1	0	0	0	0	0	4 [0.0/1.1]		
baby	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
base	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [0.0/0.0]	
bask	4	1	0	1	0	0	0	2	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	3 [0.0/0.5]
beac	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
belo	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	3	5	2	0	1	0	0	0	0	0	0	0	0	0	2	2	0	0	0	1	3	0	3 [0.0/0.6]
big	0	0	0	21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0 [87.5/0.1]
boys	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
brot	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
brot	1	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1 [88.0/0.1]	
bull	0	1	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	7 [64.7/0.1]
but	0	2	0	0	0	0	1	29	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	1	0	9 [76.3/0.2]	
cabb	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	22 [0.0/0.0]	
cons	7	2	1	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	2	0	0	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	3 [0.0/0.5]	
cost	0	1	5	0	1	0	0	1	0	0	0	0	0	1	0	0	2	5	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	2	0	0	1 [0.0/0.5]	
coup	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0 [95.2/0.0]	
cowb	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
day	1	0	1	0	1	0	0	0	0	0	0	0	0	0	2	1	4	2	2	0	0	0	0	0	4	0	0	0	0	0	0	0	0	1	2	0	2 [0.0/0.5]	
doll	14	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0 [0.0/0.6]	
doll	5	1	0	1	1	0	0	0	1	0	0	0	0	0	1	1	0	2	2	0	0	0	0	1	0	0	0	2	0	0	0	1	1	0	0	0	2 [0.0/0.5]	
drov	0	0	0	1	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1 [87.5/0.1]	
fami	5	0	0	0	0	0	0	0	0	0	0	9	0	0	0	5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	2 [42.9/0.3]	
for	0	0	1	0	0	0	0	0	0	0	0	0	14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	3 [77.8/0.1]	
girl	3	0	2	0	0	0	0	0	0	0	0	0	1	22	5	2	0	0	2	1	2	0	0	0	2	0	1	0	0	0	0	1	0	0	0	4 [50.0/0.5]		
had	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 [95.5/0.0]	
have	0	1	0	1	0	0	0	0	4	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	2	0	0	0	0	0	0	2	0	0	0	9 [0.0/0.3]		
he	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
him	3	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	10 [0.0/0.2]	
his	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	6 [83.3/0.1]	
hors	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	13	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [54.2/0.3]	
hous	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 [98.0/0.0]	
i	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	23 [96.3/0.0]		
in	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
is	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	1	1	0	0	0	0	0	0	1	0	0	0	9 [0.0/0.3]		
it	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
kans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22 [0.0/0.0]		
kids	1	0	0	2	0	0	0	0	2	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	4	3	0	0	0	0	1	1	0	0	0	7 [0.0/0.4]		
litt	3	0	1	0	0	1	0	0	0	0	1	0	0	0	4	1	1	1	7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	0	2 [0.0/0.6]		
look	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	0	0	0	12	0	0	0	1	0	0	0	0	0	0	0	0	1	0	3 [66.7/0.1]		

APÉNDICE C

PROGRAMA “prompts2mlf” para la transcripción a nivel palabra.

```
#!/usr/bin/perl

# this script makes an mlf out of a list of file names and
# corresponding prompts - ie in the format
# fileid prompt
# fileid prompt
# "      "
# The prompts are automatically converted to upper case.

if (@ARGV != 2) {
    print "usage: gettestmlf.pl mlf promptlist\n\n";
    exit (0);
}

# read in command line arguments
($mlf, $prompt) = @ARGV;

# open MLF file
open (MLF, ">$mlf") || die ("Unable to open mlf $mlf file
for writing");

    print "writing to mlf file $mlf\n";

    print MLF ("\#!MLF\!\#\n");
    # open prompt file
    open (LAB, "$prompt") || die ("Unable to open prompt
file $prompt");
    while ($line = <LAB>) {
        chop ($line);
        ($fname, @labs) = split(/\s+/, $line);
        $fname =~ s/\.mfc//g;
        $fname =~ s/\.lab//g;
        #print MLF ("\"$fname.lab\"\n");
        print MLF ("\"*\/$fname.lab\"\n");

        for ($i=0; $i<@labs; $i++) {
            #print MLF ("\U@labs[$i]\n");
            print MLF ("@labs[$i]\n");
            print MLF ("sil\n");
        }
        print MLF (".\n");
    }
close (LAB);
close (MLF);
print "writing to $mlf file done\n"
```