Rev. Téc. Ing. Univ. Zulia. Vol. 25, Nº 3, 190-201, 2002

Automatic speech recognition of Venezuelan Spanish: Test results on phrases pronounced by women

José Luciano Maldonado

Instituto de Estadística Aplicada y Computación, FACES, ULA. Núcleo La Liria, Edif. G piso 1. Mérida, Venezuela. Fax: 0274-2401116. E-mail: maldonaj@faces.ula.ve

Abstract

The objective of this work was to test the automatic recognition of phrases as pronounced by Venezuelan women. The activities included signal processing for training and testing, selection and construction of voice models, construction of the recognizer and recognition tests. The sounds were modeled using Hidden Markov Models and HTK, a publicly available toolkit. The sound files are a subset of the Venezuelan SpeechDat database. Tests were done on dates. The capacity of recognition was observed at the following levels: phonemes, words and entire phrases for each test file. To determine if the sentences were appropriate, the recognizer was equipped with a grammar which included most acceptable ways of saying dates according to Venezuelan usage. The ranges of best results were the following: at a word level, between 89.14% and 97.01%, and for entire phrases between 38.24% and 71.52%. This effort at automatic speech recognition independent of the speaker with Venezuelan voices lets us state that in the future recognizers can be built which will be able to handle the voice characteristics corresponding to this country's speech.

Key words: Speech Technology, recognition, grammar, training.

Reconocimiento automático del habla venezolana: Resultados de pruebas realizadas con frases pronunciadas por mujeres

Resumen

El objetivo de este trabajo fue hacer pruebas de reconocimiento automático de frases pronunciadas por personas de Venezuela; en particular, pronunciaciones de mujeres. Las actividades realizadas consistieron en: preparación de las señales de entrenamiento y de prueba, selección y construcción de los modelos de voz, construcción del reconocedor y pruebas de reconocimiento. Los sonidos se modelaron haciendo uso de la teoría de los Modelos Ocultos de Markov y del HTK. Las señales de voz utilizadas pertenecen a la base de datos SpeechDat Venezolana. Una vez construido el reconocedor, se le presentaron pronunciaciones de fechas y se observó su capacidad para detectar los fonos, las palabras y la fecha presente en cada pronunciación. Para determinar sí las pronunciaciones eran fechas propias del venezolano se dotó al reconocedor con una gramática con casi todas las formas en que los venezolanos pronuncian las fechas. Los rangos de reconocimiento más favorables fueron: para las palabras entre el 89.14% y el 97.01% y para las oraciones entre el 38.24% y el 71.52%. Este intento por hacer reconocimiento automático independiente del hablante con voz venezolana permite asegurar que en el futuro se podrán construir reconocedores que manejen las características propias del habla de éste País.

Palabras clave: Tecnología del habla, reconocimiento, gramática, entrenamiento.

Introducción

Uno de los subcampos de investigación de la Tecnología del Habla es el Reconocimiento Automático que tiene como objetivo darle a la máquina el don de escuchar, es decir, identificar ya sea la secuencia de fonemas, palabras y hasta oraciones presentes en una pronunciación dada como entrada, almacenar dicha secuencia y eventualmente realizar alguna acción dependiendo de la información que recibe; este subcampo, en el cual se desarrolla el presente trabajo, tiene una gran variedad de aplicaciones en la vida diaria y sin embargo, en Venezuela este tipo de tecnología ha sido muy poco explorada.

Con la idea de abordar y contribuir en algún grado con la tarea de construcción de máquinas con las características descritas y con la particularidad de manejar el habla del venezolano es que se desea mostrar los resultados de una serie de pruebas que se realizaron con pronunciaciones propias de mujeres de Venezuela.

Es conocido que en muchos centros de investigación a nivel mundial, existe una gran variedad de trabajos relacionados con el tema que se han venido realizando desde hace unos cuarenta años [1]; producto de esos trabajos hay logros importantes, sin embargo, no se ha logrado todavía construir la máquina que reciba la voz de cualquier persona, en cualquier ambiente y en cualquier lenguaje, por lo que los distintos grupos de procesamiento de la voz han enfocado sus desarrollos a incorporar poco a poco modelos de voz de unas cuantas personas, con pronunciaciones de un tema particular y en un ambiente particular con la finalidad de que en forma incremental se pueda llegar en el futuro a sistemas más generales.

En esta forma de desarrollo se enfoca el trabajo que se describe, el cual comprende la construcción de un reconocedor que recibe pronunciaciones cuya secuencia de palabras constituyen fechas tal como se pronuncian en Venezuela. El reconocimiento completo implica la identificación en secuencia de los fonemas, luego de las palabras y finalmente de las fechas.

Base de Datos

Se trabajó con archivos de voz tomados de la base de datos SpeechDat venezolana [2]. La SpeechDat Venezolana es una base de datos obtenida vía telefónica, específicamente a través de teléfonos fijos y construida bajo el formato SPEECHDAT [3]; es propiedad de la Universidad Politécnica de Cataluña, España, ya que fue obtenida como parte del Proyecto SALA de dicha Universidad, cuyo objetivo es construir una gran Base de Datos del Español hablado en el mundo y ponerla al servicio de los desarrolladores de sistemas de reconocimiento de voz para que incorporen ese lenguaje a tales sistemas.

La SpeechDat venezolana contiene grabaciones de voz de 1040 personas de las distintas regiones del país. De cada persona se obtuvieron 44 archivos, por lo que la base de datos completa consta de 45760 archivos, material suficiente para realizar estudios a nivel de ingeniería y a nivel lingüístico con el dialecto venezolano.

El contenido de la SpeechDat Venezolana comprende diversas categorías de pronunciaciones, entre las que encontramos: pronunciaciones de palabras aisladas, como dígitos, nombres, pronunciaciones de palabras conectadas como números telefónicos, números de tarjetas de crédito, pronunciaciones de frases y oraciones más complejas gramaticalmente como fechas, horas, etc. [2]. Para las pruebas de reconocimiento de las pronunciaciones de fechas descritas en este trabajo, se utilizaron 296 archivos. Entre las distintas categorías de pronunciaciones que dicha base de datos contiene, se seleccionaron archivos de fechas de mujeres de 13 estados y ciudades de Venezuela, por lo que la tarea que se describe comprende el reconocimiento automático de ese tipo de pronunciaciones.

De los 296 archivos que se utilizaron en las pruebas, 214 fueron tomados como Corpus de Entrenamiento y 82 como Corpus de Reconocimiento.

Las pronunciaciones del Corpus de Entrenamiento estaban distribuidas de la siguiente manera: 36 pronunciaciones de mujeres de Anzóategui, 13 de Sucre, 30 de Caracas, 30 de Falcón, 30 del Zulia, 15 de Portuguesa y 60 de Mérida.

Las pronunciaciones del Corpus de Reconocimiento estaban distribuidas de la siguiente manera: 4 pronunciaciones de mujeres de Barinas, 10 del Táchira, 8 de Trujillo, 6 de Bolívar, 10 de Aragua, 10 de Lara y 34 de Mérida.

Ubicación y Preparación de los Archivos de Voz

La forma como se utilizaron los archivos de voz fue la siguiente: se localizaron dentro de la base de datos SpeechDat venezolana, un conjunto de pronunciaciones de fechas de mujeres. La localización de esas señales implicaba la utilización del nombre de los archivos con los cuales aparecen en dicha bases de datos, que para el caso de fechas están reseñados con la forma A4XXXXDX.EVU donde la "D" que allí aparece, refleja que la señal corresponde a una fecha (la D es tomada de DATE del inglés), la X representa un dígito cualquiera, EVU es la extensión que indica que se trata de archivos del español hablado en Venezuela y el resto de caracteres están asociados con su codificación dentro de la base de datos del Proyecto SALA [3]. La base de datos no da más información que esa, por lo que para determinar si la voz corresponde a una mujer o a un hombre hay que tomar cada archivo y escucharlo; de la misma manera se determina el lugar de origen de las personas, puesto que para cada persona a quien se le grabó la voz hay un archivo aparte con una frase donde menciona el lugar. Esta forma de identificación, por sí sólo constituye una carga de trabajo alta.

En paralelo con la identificación descrita en el párrafo anterior, se transcribía cada pronunciación de fecha como una secuencia de símbolos, los cuales aparecen en la sección Modelos de la Voz Utilizados. Esto es lo que se ha llamado etiquetado de la voz y consistió en asociar cada sonido de la pronunciación, que se distinguía por medio del oído, con un símbolo del lenguaje español venezolano.

Posterior a ese tratamiento, cada archivo fue convertido del formato propio de la tarjeta de telefonía con que se realizaron las grabaciones a formato WAVE, con el fin de que el HTK [4] los pudiera procesar.

El último paso necesario para iniciar el modelado y la construcción del reconocedor consistió en la parametrización de las señales, para las que se obtuvieron los coeficientes cepstrales en escala de frecuencia Mel (Mel Cepstrum ó MFCC), la energía, la primera y segunda derivada, sobre segmentos de 25 milisegundos, enventanados a través de una ventana Hamming y desplazados 10 milisegundos; por lo que cada segmento se solapaba con el anterior en 15 milisegundos, además cada señal se pasaba por un filtro de pre-énfasis cuyo coeficiente tenía el valor 0.97. Se trabajó en algunas pruebas con 27 parámetros y en otras con 38.

Modelos de la Voz Utilizados

Los modelos de voz que se construyeron para el reconocimiento de las pronunciaciones de fechas están al nivel de fonos.

En los archivos utilizados en las pruebas se encontró el siguiente conjunto de fonos, al cual se ha llamado el Conjunto de Fonos Venezolanos: a, b, B, c, d, D, e, f, g, G, h, i, j, k, l, m, n, N, M, o, p, r, R, s, t, u, w, y y sil.

Se puede apreciar que no aparece la v; esto se justifica debido a que en general, en la pronunciación venezolana no se distingue la ocurrencia de una v de una b. Lo mismo sucede con la s y la z. Tampoco aparece la x, puesto que es poco frecuente el uso de tal fonema en su conversación y menos aun, en pronunciaciones de fechas. El símbolo c se seleccionó para representar el sonido de la ch y M para representar el sonido de la ñ. B, D y G representan la realización oclusiva (inicio de frase después de una nasal) de los fonemas /b/, /d/ y /g/, mientras que b, d y g representan la realización fricativa de los mismos fonemas. N representa la realización velar del fonema /n/ en distensión, es decir, al final de sílaba seguida por consonante.

Se utilizó el símbolo sil para representar las zonas de silencio presentes en las pronunciaciones. El resto de símbolos representan sonidos bien definidos e identificados en los archivos. En la sección anterior se indicó que todas las señales de voz se codificaron con éstos símbolos que representan los fonos.

Para cada uno de estos sonidos se crearon modelos con el HTK [4]. HTK es un paquete de programas que permite realizar todas las fases de procesamiento de señales para crear sistemas de reconocimiento de voz con Modelos Ocultos de Markov [1, 5].

Etiquetado de las Señales de Voz

En General, el etiquetado se realizó de la manera descrita en la sección Ubicación y Preparación de los Archivos de Voz, para el cual no era necesario alinear cada sonido con los símbolos, es decir, no se asociaba en forma manual y estrictamente cada símbolo con un segmento de la señal, ya que esto se dejaba para que HTK lo realizara automáticamente. A este proceso se le ha llamado segmentación automática.

Sin embargo, para los archivos de las fechas de las mujeres de Mérida se realizó ese tipo de etiquetado y también otra forma a la que se ha llamado segmentación semi-automática que consiste en tomar cada archivo de voz, visualizar la señal en la pantalla, luego seleccionar segmentos de esa señal y escucharlos, identificar de qué fono se trata, establecer los límites que lo separan de sus vecinos inmediatos y hacer la transcripción simbólica respectiva. En este caso, se asocia en forma directa cada fono con su respectivo símbolo, y también se establece qué segmento de la señal completa corresponde, en cuanto a tiempo de duración, a cada sonido.

No se realizó el etiquetado en forma semiautomática de todas las señales que se utilizan en el trabajo, debido a que esta actividad es demasiado lenta, tanto que por ejemplo, una persona no experta etiquetaría unas 10 señales por día, trabajando 8 horas.

Construcción de los Modelos

A partir de las 214 pronunciaciones del corpus de entrenamiento y de las realizaciones de cada sonido en dicho corpus se crearon los modelos de los fonos.

Los Modelos Ocultos de Markov que representan los fonos son del tipo Bakis [1] de 5 estados y se estimaron por medio del algoritmo Baum-Welch [1].

Inicialmente los modelos de los fonos son clones de un modelo prototipo que se creó con los vectores de medias y covarianzas globales obtenidos a partir de todos los parámetros extraídos de las señales del corpus de entrenamiento. Estos modelos inicialmente contenían una gausiana por estado.

Una vez que se obtuvieron esos modelos iniciales, el modelo de cada fono se fue re-estimando progresivamente con sus propias realizaciones hasta encontrar los mejores modelos.

El proceso de búsqueda de los mejores modelos consistió en hacer re-estimaciones y en posteriormente observar la capacidad de reconocimiento con el corpus de prueba.

A continuación se describe el proceso de re-estimación de los modelos con los cuales se obtuvieron los mejores resultados:

A partir de los modelos iniciales descritos, se realizaron 3 re-estimaciones sucesivas a través de la versión "embedded training" del algoritmo de Baum-Welch [4]. Luego se fue aumentando el número de gausianas por estado hasta llegar a seis gausianas. Cada vez que se aumentaba el número de gausianas se efectuaban dos re-estimaciones del tipo indicado antes, a excepción del último caso donde con seis gausianas por estado se realizaron entre 1 a 27 re-estimaciones más.

Diccionario

Como se mostró en la sección anterior, el sistema de reconocimiento está dotado de 29 modelos de sonidos del habla de mujeres venezolanas, lo que significa que dada una pronunciación como señal de entrada, debe ser capaz de identificar la secuencia de esos sonidos presentes en dicha pronunciación y hacer la transcripción ortográfica de dicha secuencia.

Debido a que en este trabajo nos planteamos como objetivo hacer reconocimiento de palabras y oraciones, además del reconocimiento de fonemas, dotamos al reconocedor de un diccionario cuyas entradas consistían de secuencias de sonidos (secuencia de fonos) y las salidas eran las palabras ortográficas asociadas a esas secuencias de sonidos. El diccionario se muestra en Anexos.

El diccionario contiene entonces, la lista de las palabras que intervienen en las pronunciaciones de las fechas y los sonidos asociados a esas palabras.

En el diccionario se puede observar una misma palabra asociada con secuencias diferentes de sonidos, esto se debe a que se trató de cubrir para el reconocimiento, las diferentes va-

riantes de pronunciación de las palabras que se detectaron en los archivos bajo estudio.

Para ilustrar esta situación, presentamos como ejemplo la palabra diciembre, que puede estar constituida por la secuencias de sonidos siguientes: disjembre, Disjembre, disjemBre ó DisjemBre.

Esto significa que como las personas pronuncian las mismas palabras de manera diferente, entonces hay que buscar una forma adecuada para asociar diferentes secuencias de sonidos a una misma palabra. Esto se logra construyendo el diccionario en la manera descrita.

Gramática

Así como a través del diccionario se puede identificar la secuencia de palabras presente en una pronunciación, para identificar oraciones hay que dotar a los reconocedores con las reglas gramaticales del lenguaje al cual están asociadas las pronunciaciones que se pretenda que éstos manejen. En nuestro caso, se incorporó al reconocedor una gramática que consiste en la combinación adecuada de todas las palabras que usan los venezolanos para pronunciar o escribir fechas. En Anexos aparece la programación de esa gramática, la cual está escrita en el formato HTK.

El programa que modela la gramática permite aceptar fechas de las formas que mostramos a continuación como ejemplo:

"El martes cuatro de diciembre de mil novecientos treinta y cuatro", "martes cuatro de diciembre de mil novecientos treinta y cuatro", "cuatro de diciembre de mil novecientos treinta y cuatro", "diciembre de mil novecientos treinta y cuatro", "cuatro doce de mil novecientos treinta y cuatro", "cuatro doce treinta y cuatro", "el cuatro doce de mil novecientos treinta y cuatro", "en el mes de diciembre de mil novecientos treinta y cuatro", "en diciembre del treinta y cuatro" y algunas otras.

Hay un tipo de realización de fechas que consiste en una secuencia de números como: "cero cuatro cero cinco cero nueve", que no se consideró en nuestro experimento, y que se tomará en cuenta en futuras pruebas.

Descripción de las Pruebas Realizadas

A continuación se describen las pruebas realizadas: en primer lugar, se describe la prueba de reconocimiento de fechas de las mujeres de Mérida, que constituyó el punto de arranque para la segunda prueba, que es más general, y que consistió en el reconocimiento de fechas de mujeres de Venezuela.

Pruebas con voces de mujeres de Mérida

Estas pruebas se realizaron a partir de reconocedores construidos en base a 60 archivos de entrenamiento y a 34 archivos del corpus de reconocimiento.

En este reconocimiento se realizaron dos tipos de pruebas: En el primer caso, la segmentación de la duración de los sonidos se hizo en forma semi-automática y en el segundo caso, dicha segmentación se realizó en forma automática. Para el primer tipo de pruebas, el mejor reconocedor fue construido en base a 15 re-estimaciones por modelo, donde se trabajó con 27 parámetros por frame de la señal y para el segundo tipo de pruebas, el mejor reconocedor fue construido en base a 54 re-estimaciones por modelo y 38 parámetros por frame.

Pruebas con voces de mujeres de Venezuela

Estas pruebas se realizaron a partir de reconocedores construidos en base a todos los archivos del corpus de entrenamiento y a 48 archivos del corpus de reconocimiento (no se utilizaron los 34 archivos de mujeres de Mérida de éste último corpus).

El etiquetado y la segmentación de los sonidos se hizo en forma automática. Se trabajó con 38 parámetros por frame y el mejor reconocedor fue construido en base a 39 re-estimaciones por modelo.

En este caso se realizaron pruebas separadas donde se utilizaron archivos que formaban parte del corpus de entrenamiento, y pruebas con archivos que formaban parte del corpus de

reconocimiento. Los archivos para estas últimas pruebas fueron tomados de personas con orígenes distintos a los orígenes de las personas del entrenamiento.

Para este tipo de reconocimiento se trabajó en un caso con la transcripción general para el habla Venezolana llamada Conjunto de Fonos Venezolanos (ver sección Modelos de la Voz Utilizados), y en otro caso, con transcripciones alternativas producto de las pronunciaciones que se han venido escuchando en la SpeechDat venezolana y de nuestra experiencia diaria.

Todas las pruebas de reconocimiento se realizaron con el algoritmo Viterbi del HTK [1].

Resultados

Resultados más favorables en las pruebas de reconocimiento de fechas de las mujeres de Mérida

Para la primera prueba, la capacidad de reconocimiento resultó del 92.57% a nivel de palabras (324 palabras reconocidas correctamente de las 350 que intervienen en la prueba) y del 38.24% a nivel de fechas (13 fechas correctas de las 34 que intervienen).

En los resultados de la segunda prueba encontramos una capacidad de reconocimiento del 89.14% a nivel de palabras (312 palabras reconocidas correctamente de las 350 que intervienen en la prueba) y del 41.18% a nivel de fechas (14 fechas correctas de las 34 que intervienen en la prueba).

Resultados más favorables en las pruebas de reconocimiento de fechas de las mujeres de Venezuela

Para el caso donde se utilizaron transcripciones del Conjunto de Fonos Venezolanos y con 151 pronunciaciones tomadas del corpus de entrenamiento se encontró una capacidad de reconocimiento del 96.94% a nivel de palabras (1362 palabras reconocidas correctamente de las 1405 que intervienen en la prueba) y del 70.86% a nivel de fechas (107 fechas correctas de las 151 que intervienen en la prueba).

Para el caso donde se utilizaron transcripciones del Conjunto de Fonos Venezolanos y con 38 pronunciaciones tomadas del corpus de reconocimiento (después de revisar cuidadosamente las 48 pronunciaciones restantes sin tomar en cuenta las de Mérida, se encontraron archivos que contenian algunas palabras que normalmente no están presentes en las fechas y otros que tenian un nivel de ruido alto, por lo que fueron descartados 10 de esos archivos) se encontró una capacidad de reconocimiento del 95.88% a nivel de palabras (349 palabras reconocidas correctamente de las 364 que intervienen en la prueba) y del 63.16% a nivel de fechas (24 fechas correctas de las 38 que intervienen en la prueba).

Para el caso donde se utilizaron transcripciones alternativas y con las 151 pronunciaciones tomadas del corpus de entrenamiento se encontró una capacidad de reconocimiento del 97.01% a nivel de palabras (1363 palabras reconocidas correctamente de las 1405 que intervienen en la prueba) y del 71.52% a nivel de fechas (108 fechas correctas de las 151 que intervienen en la prueba).

Para el caso donde se utilizaron transcripciones alternativas y las 38 pronunciaciones de reconocimiento descritas se encontró una capacidad de reconocimiento del 96.98% a nivel de palabras (353 palabras reconocidas correctamente de las 364 que intervienen en la prueba) y del 60.53% a nivel de fechas (23 fechas correctas de las 38 que intervienen en la prueba).

En Anexos se muestra en detalle el resultado de la última prueba y se resume los resultados de las restantes.

Discusión de Resultados

Los resultados alcanzados son satisfactorios, especialmente si se toma en cuenta que no sólo se estaba haciendo reconocimiento a nivel de fonos y de palabras, sino también de oraciones, y que para lograr este último tipo de reconocimiento se requieren buenos resultados a nivel de las unidades previas, y eso sólo se logra cuando los modelos están bien entrenados. Otro aspecto a considerar a la hora de evaluar los resultados, es el hecho de que se trabajó con voces de múltiples hablantes lo que le da amplitud a los reconocedores, puesto que no es esclavo de la voz de una sola persona.

Para el caso del reconocimiento de fechas con voces de mujeres de Mérida, el resultado obtenido, cuando se utiliza etiquetado y segmentación manual, es mejor en 3.43% a nivel de palabras, mientras que en el reconocimiento de frases es menor en un 2.94%. Esto último, podría suponer una contradicción, que no es tal debido a que así como hay frases en las que se disminuye la identificación correcta de algunas palabras, en otras puede aumentarse y por lo tanto, algunas oraciones que antes resultaron no completamente reconocidas es posible que posteriormente puedan ser reconocidas en su totalidad.

Para el caso general, donde se hace reconocimiento de fechas con voces de mujeres de buena parte del territorio venezolano, se puede observar que la capacidad de reconocimiento, cuando se trabaja con algunas transcripciones alternativas para los sonidos, es 1.1% mejor a nivel de palabras al reconocimiento del caso cuando se trabaja con las transcripciones generales. Sin embargo, a nivel de frases completas, el nivel de reconocimiento con transcripciones generales, resulta mejor en 2.63%; este es el caso en que se trabajó con el corpus de reconocimiento, mientras que con el grupo de frases correspondientes al corpus de entrenamiento, el resultado estuvo siempre a favor del caso de transcripciones alternativas: 0.07% mejor a nivel de palabras y 0.66% mejor a nivel de frases. Desde luego, la contradicción aparente que se presenta cuando se trabaja con el corpus de reconocimiento, se explica como en el párrafo anterior.

De los resultados globales de las pruebas, se puede observar que el nivel de reconocimiento con las voces de mujeres de Venezuela es superior al obtenido cuando se trabajó con voces de las mujeres de Mérida, lo que es una señal clara de que los modelos están mejor entrenados. También, el etiquetado y la segmentación semi-automática, así como el trabajo con transcripciones alternativas generan mayor capacidad de reconocimiento a nivel de palabras, lo que hace suponer que en general, el reconocimiento a nivel de frases también será superior.

Por otro lado, se puede manifestar que los resultados obtenidos están en el rango cuantitativo obtenido por diversos grupos a nivel mundial en experimentos donde se hace reconocimiento de voz continua como es este caso [5, 6].

Conclusiones

Es posible crear modelos del habla de mujeres venezolanas a nivel de fonos para hacer reconocimiento automático de oraciones.

No se justifica el etiquetado y la segmentación manual si hay herramientas para hacerlo automáticamente lo que genera menos costo de trabajo, aun cuando los resultados que se obtengan puedan ser relativamente superiores..

Es posible crear reconocedores generales del habla de mujeres venezolanas, que puedan admitir la voz de muchas mujeres y hasta ser independientes del hablante.

Los resultados obtenidos permiten suponer que en el futuro se podrán construir reconocedores del español hablado en Venezuela.

Se requiere una gran cantidad de pronunciaciones de entrenamiento por modelo para obtener una capacidad de reconocimiento alta.

Referencias Bibliográficas

1. Deller J., Proakis J. y Jansen J., Discrete-Time Processing of Speech Signals. New York: MacMillan Publishing Company 1993.
2. Moreno A., Mora E., Speechdat Spanish Venezuelan database for the fixed Telephone network, Universidad politécnica de Cataluña, España y Universidad de Los Andes, Venezuela. 1999.
3. Moreno A., Speechdat Spanish Database for the fixed Telephone Network. Universidad Politécnica de Cataluña, España. 1997.
4. Entropic Speech Technology, HTK Manual.
5. Zhao Y., A Speaker-Independent Continuous Speech Recognition System using Continuous mixture Gaussian Density HMM of phoneme-sized units. IEEE Transaction on Speech and Audio Processing, Vol. 1, No. 3, July (1993).
6. Savage J., A hybrid System with Symbolic AI and Statistical Methods for Speech Recognition. A dissertation for the degree of Doctor of Philosophy, University of Washington, USA 1995.

Recibido el 18 de Septiembre de 2001
En forma revisada el 02 de Junio de 2002

Anexos

Diccionario

abril a b r i l
agosto a g o s t o
agosto a g o h t o
año a M o
catorce k a t o r s e
cero s e r o
cinco s i N k o
cincuenta s i N k w e N t a
cuarenta k w a r e N t a
cuatro k w a t r o
de d e
de D e
del d e l
del D e l
diciembre d i s j e m b r e
diciembre D i s j e m b r e
diciembre d i s j e m B r e
diciembre D i s j e m B r e
diecinueve d j e s i n w e b e
diecinueve D j e s i n w e b e
dieciocho d j e s i o c o
dieciocho D j e s i o c o
dieciseis d j e s i s e j s
dieciseis d j e s i s e j
dieciseis D j e s i s e j s
dieciseis D j e s i s e j
diecisiete d j e s i s j e t e
diecisiete D j e s i s j e t e
diez d j e s
diez d j e
diez D j e s
diez D j e
diez d j e h
diez D j e h
doce d o s e
doce D o s e
domingo d o m i N G o
domingo d o m i G o
domingo D o m i N G o
domingo D o m i G o
dos d o s
dos d o
dos D o s
dos D o
dos d o h
dos D o h
el e l
en e n
enero e n e r o
febrero f e b r e r o
jueves h w e b e s
jueves h w e b e
jueves h w e b e h
julio h u l j o
junio h u n j o
lunes l u n e s
lunes l u n e h
lunes l u n e
martes m a r t e s
martes m a r t e
martes m a r t e h
marzo m a r s o
mayo m a y o
mes m e s
mes m e h
miercoles m j e r k o l e s
miercoles m j e r k o l e
miercoles m j e r k o l e h
mil m i l
mil m i
novecientos n o b e s j e N t o s
novecientos n o b e s j e N t o
novecientos n o b e s j e N t o h
noventa n o b e N t a
noventa n o b e N t
noviembre n o b j e m b r e
noviembre n o b j e m B r e
nueve n w e b e
ochenta o c e N t a
ocho o c o
octubre o t u b r e
once o n s e
proximo p r o s i m o
quince k i n s e
sabado s a b a d o
seis s e j s
seis s e j
septiembre s e t j e m b r e
septiembre s e t j e m B r e
sesenta s e s e N t a
setenta s e t e N t a
setenta s e t e N t
siete s j e t e
trece t r e s e
treinta t r e j N t a
tres t r e s
tres t r e
tres t r e h

```
uno u n o
veinte b e j N t e
veinte B e j N t e
veinticinco B e j N t i s i N k o
veinticinco b e j N t i s i N k o
veinticuatro b e j N t i k w a t r o
veinticuatro B e j N t i k w a t r o
veintidos b e j N t i d o s
veintidos B e j N t i d o
veintidos b e j N t i d o h
veintidos B e j N t i d o h
veintidos B e j N t i d o s
veintidos b e j N t i d o
veintidos b e j N t i o s
veintidos B e j N t i o
veintidos B e j N t i o h
veintinueve b e j N t i n w e b e
veintinueve B e j N t i n w e b e
veintiocho b e j N t i o c o
veintiocho b e N t i o c
veintiocho B e j N t i o c o
veintiseis b e j N t i s e j s
veintiseis b e j N t i s e j
veintiseis B e j N t i s e j s
veintiseis b e j N t i s e j h
veintiseis B e j N t i s e j h
veintiseis B e j N t i s e j
veintisiete b e j N t i s j e t e
veintisiete B e j N t i s j e t e
veintitres b e j N t i t r e s
veintitres b e j N t i t r e
veintitres B e j N t i t r e s
veintitres b e j N t i t r e h
veintitres B e j N t i t r e h
veintitres B e j N t i t r e
veintiuno b e j N t i u n o
veintiuno B e j N t i u n o
viernes b j e r n e s
viernes b j e r n e
viernes B j e r n e s
viernes b j e r n e h
viernes B j e r n e h
viernes B j e r n e
y i
y y
SIL SIL
```

Gramática

```
$num1 = cero | uno | dos | tres | cuatro |
        cinco | seis | siete | ocho | nueve;
$num2 = diez | once | doce | trece | catorce |
        quince | dieciseis | diecisiete | dieciocho
        | diecinueve | veinte | veintiuno |
        veintidos | veintitres | veinticuatro |
        veinticinco | veintiseis | veintisiete |
        veintiocho | veintinueve | treinta;
$num3 = cuarenta | cincuenta | sesenta |
        setenta | ochenta | noventa;
$fin2 = y $num1;
$fin1 = $num1 | $num2 | $num3 $fin2 |
        $num3;
$nmes = $num1 | diez | once | doce;
$dia = lunes | martes | miércoles | jueves |
        viernes | sábado | domingo;
$mes = enero | febrero | marzo | abril | mayo
        | junio | julio | agosto | septiembre |
        octubre | noviembre | diciembre;
$delanodosmil = del año dos mil;
$deldosmil = del dos mil;
$delanomil = del año mil;
$demil = de [SP] mil | del [SP] mil;
$dosmil = $delanodosmil | $deldosmil;
$ano = $delanomil | $demil;
$demes = de $mes | $mes;
$eldia = el $dia | $dia;
$enmes = en $mes;
$num = $num1 | $num2;
$elnum = el $num | $num;
$parte4 = $eldia [SP] $num | el [SP] $num |
        $num;
$parte1 = $parte4 $demes | $enmes | $elnum
        $nmes;
$parte2 = $ano | $dosmil;
$parte3 = $fin1;
$parte5 = $parte2 novecientos | $parte2;
$parte6 = $parte5 $parte3 | del $parte3;
$parte7 = $parte4 $demes;
( SIL ($parte7 | $parte1 [SP] $parte6 ) SIL)
```

Resultados de la Primera Prueba

Overall Results

```
SENT: %Correct=38.24 [H=13, S=21, N=34]
WORD: %Corr=92.57, Acc=85.14 [H=324, D=5,
      S=21, I=26, N=350]
==================================
```

Resultados de la Segunda Prueba

Overall Results

```
SENT: %Correct=41.18 [H=14, S=20, N=34]
WORD: %Corr=89.14, Acc=84.86 [H=312, D=9,
      S=29, I=15, N=350]
```

==========

Resultados de la Tercera Prueba
Overall Results
SENT: %Correct=70.86 [H=107, S=44, N=151]
WORD: %Corr=96.94, Acc=95.37 [H=1362, D=13, S=30, I=22, N=1405]
==========

Resultados de la Cuarta Prueba
Overall Results
SENT: %Correct=63.16 [H=24, S=14, N=38]
WORD: %Corr=95.88, Acc=93.96 [H=349, D=3, S=12, I=7, N=364]
==========

Resultados de la Quinta Prueba
Overall Results
SENT: %Correct=71.52 [H=108, S=43, N=151]
WORD: %Corr=97.01, Acc=95.30 [H=1363, D=12, S=30, I=24, N=1405]
==========

Resultados de la Sexta Prueba
HTK Results Analysis
Date: Tue Feb 26 09:20:55 2002
Ref: /home/luciano/tesisfechas/listas/palabrasmujerestest.mlf
Rec: /home/luciano/tesisfechas/reconocidos/MLFsalida

File Results
a40002d1.rec: 100.00(100.00) [H= 11, D= 0, S= 0, I= 0, N= 11]
a40012d2.rec: 88.89(88.89) [H= 8, D= 0, S= 1, I= 0, N= 9]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40012d2.lab vs /home/luciano/tesisfechas/reconocidos/a40012d2.rec
LAB: sil martes miercoles seis de abril de dos mil veintinueve
REC: sil martes miercoles seis de abril del dos mil veintinueve
a40012d3.rec: 100.00(83.33) [H= 6, D= 0, S= 0, I= 1, N= 6]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40012d3.lab vs /home/luciano/tesisfechas/reconocidos/a40012d3.rec
LAB: sil el veintinueve de mayo sil
REC: sil el veintinueve de mayo cinco sil
a40020d1.rec: 100.00(100.00) [H= 11, D= 0, S= 0, I= 0, N= 11]
a40037d1.rec: 90.91(90.91) [H= 10, D= 0, S= 1, I= 0, N= 11]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40037d1.lab vs /home/luciano/tesisfechas/reconocidos/a40037d1.rec
LAB: el cuatro de agosto de mil novecientos ochenta y uno sil
REC: jueves cuatro de agosto de mil novecientos ochenta y uno sil
a40037d3.rec: 100.00(100.00) [H= 6, D= 0, S= 0, I= 0, N= 6]
a40054d2.rec: 100.00(100.00) [H= 12, D= 0, S= 0, I= 0, N= 12]
a40060d1.rec: 100.00(90.91) [H= 11, D= 0, S= 0, I= 1, N= 11]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40060d1.lab vs /home/luciano/tesisfechas/reconocidos/a40060d1.rec
LAB: sil dos de junio de mil novecientos setenta y tres sil
REC: sil sil dos de junio de mil novecientos setenta y tres sil
a40060d2.rec: 100.00(100.00) [H= 10, D= 0, S= 0, I= 0, N= 10]
a40060d3.rec: 83.33(83.33) [H= 5, D= 0, S= 1, I= 0, N= 6]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40060d3.lab vs /home/luciano/tesisfechas/reconocidos/a40060d3.rec
LAB: sil el diecisiete de mayo sil
REC: sil el diecisiete de marzo sil
a40066d2.rec: 90.91(90.91) [H= 10, D= 0, S= 1, I= 0, N= 11]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40066d2.lab vs /home/luciano/tesisfechas/reconocidos/a40066d2.rec

LAB: sil martes seis de agosto de mil novecientos cincuenta y uno
REC: sil martes seis de agosto de mil novecientos cincuenta y dos
a40067d2.rec: 100.00(100.00) [H= 13, D= 0, S= 0, I= 0, N= 13]
a40070d2.rec: 100.00(100.00) [H= 9, D= 0, S= 0, I= 0, N= 9]
a40081d2.rec: 100.00(100.00) [H= 12, D= 0, S= 0, I= 0, N= 12]
a40081d3.rec: 85.71(85.71) [H= 6, D= 0, S= 1, I= 0, N= 7]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40081d3.lab vs /home/luciano/tesisfechas/reconocidos/a40081d3.rec
LAB: sil el proximo mes de mayo sil
REC: sil el proximo mes de marzo sil
a40085d1.rec: 87.50(87.50) [H= 7, D= 0, S= 1, I= 0, N= 8]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40085d1.lab vs /home/luciano/tesisfechas/reconocidos/a40085d1.rec
LAB: sil diez ocho del ochenta y uno sil
REC: sil dieciocho ocho del ochenta y uno sil
a40085d2.rec: 100.00(100.00) [H= 14, D= 0, S= 0, I= 0, N= 14]
a40107d3.rec: 100.00(100.00) [H= 5, D= 0, S= 0, I= 0, N= 5]
a40139d2.rec: 100.00(100.00) [H= 11, D= 0, S= 0, I= 0, N= 11]
a40139d3.rec: 83.33(66.67) [H= 5, D= 0, S= 1, I= 1, N= 6]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40139d3.lab vs /home/luciano/tesisfechas/reconocidos/a40139d3.rec
LAB: el proximo mes de mayo sil
REC: sil el proximo mes de marzo sil
a40146d2.rec: 100.00(100.00) [H= 13, D= 0, S= 0, I= 0, N= 13]
a40146d3.rec: 100.00(100.00) [H= 7, D= 0, S= 0, I= 0, N= 7]
a40176d1.rec: 100.00(100.00) [H= 11, D= 0, S= 0, I= 0, N= 11]
a40176d2.rec: 90.91(81.82) [H= 10, D= 0, S= 1, I= 1, N= 11]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40176d2.lab vs /home/luciano/tesisfechas/reconocidos/a40176d2.rec
LAB: sil viernes sil veinticuatro de septiembre de mil sil quince sil
REC: sil viernes sil veinticuatro de septiembre de mil sil y seis sil
a40176d3.rec: 100.00(100.00) [H= 6, D= 0, S= 0, I= 0, N= 6]
a40184d1.rec: 100.00(100.00) [H= 9, D= 0, S= 0, I= 0, N= 9]
a40184d2.rec: 100.00(100.00) [H= 8, D= 0, S= 0, I= 0, N= 8]
a40184d3.rec: 100.00(60.00) [H= 5, D= 0, S= 0, I= 2, N= 5]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40184d3.lab vs /home/luciano/tesisfechas/reconocidos/a40184d3.rec
LAB: el tres de marzo sil
REC: sil el sil tres de marzo sil
a40257d2.rec: 100.00(100.00) [H= 10, D= 0, S= 0, I= 0, N= 10]
a40285d3.rec: 100.00(100.00) [H= 5, D= 0, S= 0, I= 0, N= 5]
a40301d1.rec: 100.00(100.00) [H= 11, D= 0, S= 0, I= 0, N= 11]
a40302d2.rec: 100.00(100.00) [H= 13, D= 0, S= 0, I= 0, N= 13]
a40407d2.rec: 90.91(90.91) [H= 10, D= 0, S= 1, I= 0, N= 11]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40407d2.lab vs /home/luciano/tesisfechas/reconocidos/a40407d2.rec
LAB: sil martes veinticuatro de mayo sil del dos mil cuatro sil
REC: sil martes veinticuatro de mayo sil del dos mil dos sil
a40433d1.rec: 100.00(100.00) [H= 11, D= 0, S= 0, I= 0, N= 11]
a40433d2.rec: 100.00(90.00) [H= 10, D= 0, S= 0, I= 1, N= 10]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40433d2.lab vs

/home/luciano/tesisfechas/reconocidos/a40433d2.rec
LAB: sil miercoles catorce de mayo del dos mil veinticuatro sil
REC: sil el miercoles catorce de mayo del dos mil veinticuatro sil
a40476d2.rec: 90.91(90.91) [H= 10, D= 1, S= 0, I= 0, N= 11]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40476d2.lab vs
/home/luciano/tesisfechas/reconocidos/a40476d2.rec
LAB: sil domingo veintiocho de septiembre de del dos mil diecinueve sil
REC: sil domingo veintiocho de septiembre del dos mil diecinueve sil

a40479d1.rec: 100.00(100.00) [H= 11, D= 0, S= 0, I= 0, N= 11]
a40479d2.rec: 91.67(91.67) [H= 11, D= 0, S= 1, I= 0, N= 12]
Aligned transcription:
/home/luciano/etiquetas/mujeres/fechas/test/palabras/a40479d2.lab vs
/home/luciano/tesisfechas/reconocidos/a40479d2.rec
LAB: martes sil veintiocho de febrero de mil novecientos sesenta y seis sil
REC: martes sil veintiocho de febrero de mil novecientos setenta y seis sil

Overall Results
SENT: %Correct=60.53 [H=23, S=15, N=38]
WORD: %Corr=96.98, Acc=95.05 [H=353, D=1, S=10, I=7, N=364]
==========