

Diseño de un Corpus de Voz en Español para Niños en Edad Escolar con Problemas de Lenguaje

Glendy Perera-Góngora
y Carlos Miranda-Palma

Resumen

En este trabajo se describe la creación de un corpus de voz para su uso en niños en edad escolar con problemas de lenguaje. Se describen los elementos del corpus de voz, su implementación en un juego electrónico con reconocimiento de voz y los resultados obtenidos en niños con y sin problemas de lenguaje. Para el desarrollo de este trabajo se utilizó CSLU ToolKit, la herramienta HTK (Hidden ToolKit), el Lenguaje C y las librerías SDL.



I. Introducción

En estos tiempos, en donde la tecnología se ha hecho indispensable para el desarrollo de las actividades laborales, personales, profesionales, de entretenimiento, estudio, la computadora es el instrumento ideal para realizar diversas funciones, entre las que se destacan trabajos sobre la ingeniería de software, paqueterías, creación de gráficos, programas interactivos para la enseñanza, herramientas para la robótica, visión artificial, el reconocimiento de voz, etc. En particular, en el área de reconocimiento de voz se ha logrado cierto avance, pues en los trabajos realizados hasta ahora, supone un alto nivel de reconocimiento de las palabras, por lo que los resultados han sido favorables en cada uno de los enfoques diseñados.

Dada la importancia de esta área de la inteligencia artificial, es importante hacer alusión a los beneficios aportados al sistema educativo, especialmente en los niveles de atención para personas con algún tipo de discapacidad, ya que la forma novedosa en que se presenta, permite despertar el interés y la necesidad de utilizar esta área para el máximo aprovechamiento de la computadora.

Ejemplos de estas aportaciones son los sistemas de reconocimiento de voz que se han creado para las diversas edades y lenguajes del ser humano. Como elemento fundamental de un sistema de reconocimiento de voz existe el Corpus de voz, el cual se define como un conjunto de grabaciones con formato de audio, así como un conjunto de comentarios y documentos asociados a los datos de voz. Dichos datos deben estar almacenados con un formato estándar para ser utilizados por el equipo de cómputo (Luna, 2001).

En el Instituto de Investigaciones en Matemáticas aplicadas y Sistemas, se desarrolló un trabajo, el cual pretendía el estudio de la interacción multimodal hombre – máquina, donde a través de la plataforma, un ser humano simulaba las capacidades de entendimiento y comprensión del sistema e interactuaba con el usuario a través de una interfaz computacional real (Villaseñor, 2000). Asimismo, en esta misma área se han creado algunos corpus de voz para el español, entre ellos está el DIME (Villaseñor, 2001) elaborado en el IIMAS, el CORDE (RAE, 2008) realizado por la Real Academia Española entre otros, que han servido como apoyo para diversas funciones, desde entretenimiento hasta la enseñanza. En el ITESM Campus Cuernavaca, se diseñó un corpus de voz para manipular un robot móvil evitando obstáculos por medio de la voz. Este corpus no sólo permitía reconocer los comandos que ejecuta el robot Noman Scout II, sino que puede reconocer otras palabras. El modelado fue realizado utilizando el método de los Modelos Ocultos de Markov (Miranda, 2003).

También existen trabajos que incorporan un corpus de voz para niños. Tal es el caso del corpus realizado en la Universidad de las Américas, Puebla (Guzmán, 2004), cuyo objetivo fue desarrollar un sintetizador de voz (proceso de transformar el texto en sonido), que comprende sonidos específicos para una pronunciación ideal en el contexto de la enseñanza para los niños mexicanos.

Otro corpus está enfocado hacia la realización de un sistema tutorial para la enseñanza del vocabulario hablado en México a niños de primaria, utilizando tecnología de reconocimiento y síntesis de voz (Luna, 2001). El objetivo principal de ese trabajo fue la elaboración de un sistema que enseñe a niños el lenguaje español hablado en México mediante la tecnología del reconocimiento de voz. El modelado fue realizado usando redes neuronales.

En la Universidad de Valencia, España, se realizó un estudio, en el cual se compararon dos sistemas de reconocimiento de voz de habla discreta en personas adultas con problemas del habla causado por diferentes problemas, como parálisis cerebral, entre otros. Los resultados muestran que ambos sistemas alcanzaron niveles de reconocimiento correcto superiores al 90% (Avila, 2001).

La familia, los amigos, la escuela, etc., son quienes influyen en nuestra capacidad de hablar, de manera particular en el desenvolvimiento futuro, pues es nuestro medio de comunicación; sin embargo, para las personas -en particular los niños- con trastornos del habla, es de mayor importancia, debido a sus dificultades de expresión, que interfieren y limitan su conducta de comunicación con los demás y su comportamiento de adaptación y ajuste al medio (Castañeda, 1999).

En la actualidad en México hay una gran carencia en software de reconocimiento de voz orientado a la educación para niños; y el que existe, es de uso general, dejando a un lado a las personas con necesidades especiales.

En este trabajo se presenta el desarrollo de un sistema con reconocimiento de voz que sea un auxiliar en la mejora de la habilidad oral en niños con problemas del habla en idioma español, así como un avance en el sistema de reconocimiento de voz, proporcionando información sobre los resultados obtenidos de una comparación entre niños con y sin problemas de lenguaje.

II. Características de los hablantes

Para este trabajo, el corpus fue grabado con un grupo mediano de hablantes (42 niños) de tres escuelas primarias de la ciudad de Tizimín, Yucatán y el tipo de hablantes son hombres y mujeres de distinta edad (primer a sexto grado de primaria y entre 7 y 13 años de edad).

En la tabla I se muestra las características de los hablantes por escuela y grado.

TABLA I
GRADO ESCOLAR DE LOS HABLANTES

Grado / Escuela	Escuela 1	Escuela 2	Escuela 3
Primer	3	2	2
Segundo	3	2	2
Tercer	3	2	2
Cuarto	3	2	2
Quinto	3	2	2
Sexto	3	2	2

III. Interfaz gráfica

La idea de realizar una interfaz gráfica nació de la necesidad de las maestras especializadas en el área de problemas de lenguaje de contar con nuevas herramientas en su trabajo de rehabilitación, ya que de manera habitual se basan en la repetición de palabras, observando tarjetas o imágenes completamente estáticas, métodos que resultan ser agotadores e incluso fastidiosos en algunas ocasiones.

La técnica para corregir el problema del habla (en particular la dislalia) es la repetición. Esta misma técnica es la usada en este trabajo, con la variante de que un videojuego, además de interactivo, es llamativo para los niños.

La realización de la interfaz gráfica consistió en crear un ambiente de competencia con la computadora, lo cual favorece el propósito de este trabajo, ya que ayuda en la práctica por medio de la repetición a los niños con problemas de lenguaje.



Fig. 3. Interfaz gráfica con reconocimiento de voz



Fig. 4. Interfaz gráfica con reconocimiento de voz

el personaje contrario (que representa a la computadora) es quien avanza.

El juego se termina cuando uno de los dos personajes llega a la meta o cuando se pronuncia las palabras "SALIR" o "TERMINAR" emitiendo un mensaje de felicitación o de ánimo para repetir (Fig 4 y Fig 5).

Como se puede observar, el juego es más una competencia con la computadora y hasta con el niño mismo, pues el objetivo es pronunciar de manera correcta la palabra.



Fig. 5. Interfaz gráfica con reconocimiento de voz

IV. Pruebas

La metodología empleada para evaluar el desempeño del sistema de reconocimiento de voz se basa en el criterio de la evaluación del porcentaje de palabras reconocidas correctamente. Este porcentaje se determina a partir de las palabras no reconocidas por el sistema.

$$TR = 35 - \text{no_reconocidas.}$$

$$\text{Porcentaje} = (TR * 100) / 35$$

Donde:

Porcentaje, es la cantidad correspondiente al porcentaje de palabras reconocidas de manera correcta por el reconocedor.

no_reconocidas, es la cantidad correspondiente a las palabras no reconocidas por el sistema

TR, es la cantidad de palabras reconocidas.

Las pruebas del sistema desarrollado se realizaron con 7 niños de diversos grados de primaria de la ciudad de Tizimín, Yucatán, de entre 8 y 13 años de edad, que fueron elegidos al azar. La evaluación se realizó mediante la observación del desempeño de los niños en el desarrollo del videojuego, anotando el número de aciertos, así como las palabras que no reconocían. Es importante hacer notar que de este grupo de niños, se puede clasificar en dos grupos: niños con problemas del habla y niños sin problemas de habla. Los niños nombrados 1, 2 y 3 corresponden al primer grupo, es decir, niños con algún problema de lenguaje (ConPL) y los cuatro restantes, son niños que no tienen problema con el lenguaje (SinPL).

TABLA II
RESULTADOS DE LAS PRUEBAS

En la tabla II se presentan los resultados de la evaluación. Se puede observar que el rango de porcentaje de reconocimiento no es muy distante entre los dos grupos de hablantes.

	Niño		Palabras	
	ConPL	SinPL	No reconocidas	Porcentaje
1			5	95
2			9	90
3			7	93
		4	2	98
		5	3	97
		6	0	100
		7	2	98

V. Desarrollo del corpus

En esta sección se describe el concepto de corpus de voz, el proceso para desarrollar un corpus de voz y las etapas del entrenamiento de los modelos fonéticos.

A. Definición

Como se mencionó anteriormente, un corpus de voz es un conjunto de grabaciones con formato de audio, así como un conjunto de comentarios y documentos a los datos de voz. Dichos datos deben estar almacenados con un formato estándar para ser utilizados por el equipo de cómputo (Luna, 2001).

Las etapas para desarrollar el corpus de voz fueron las siguientes:

Diseño: En esta etapa se determinó el contenido del corpus, es decir, se crearon las frases que posteriormente fueron utilizadas para la etapa de grabación. Las frases fueron seleccionadas de libros de español utilizados en primaria.

Grabación: En esta etapa se realizaron las grabaciones de voz. El trabajo consistió en pedir a un grupo de niños seleccionados que leyeran de manera natural las frases creadas en la etapa anterior. Para esta etapa se utilizó la herramienta record.tk del CSLU Toolkit, que permite leer una frase, almacenarla y revisarla.

La instrucción para utilizar esta herramienta es la siguiente:

```
\progra~1\cslu\tcl80\bin\wish80 record.tk -base /rad/data -user 0 -prompts <nom_archivo.txt>
```

Transcripciones: En esta etapa se realizó la creación de los modelos fonéticos; para esto se utilizó HTK (Hidden ToolKit). Esta etapa se dividió en 4 fases (Preparación de datos, Entrenamiento, Prueba y Análisis).

B. Creación de los Modelos Fonéticos

El proceso para la creación de los modelos fonéticos está basado en la preparación de datos del corpus Voxforge (Voxforge, 2008) y apoyado por el HTK Book (HTK, 2008). A continuación se describe el proceso para la etapa de transcripción de datos, las cuales se dividen en:

1) Fase de Preparación de Datos

Una vez finalizadas las grabaciones de las frases definidas en la etapa de grabación, se creó un archivo de texto que contiene los datos de las frases, es decir, contiene los nombres de los archivos de texto seguidos de la frase, eliminando los signos de puntuación, los caracteres extraños y convirtiendo las letras a mayúsculas, es decir, se realizó una transcripción a nivel ortográfico. Utilizando este archivo, se generó uno nuevo que contiene la lista de las palabras usadas en las frases.

Asimismo, se creó un archivo de texto que contiene el universo de palabras, acompañado del formato en el cual éstas se mostrarán en la pantalla y los caracteres que representan

su pronunciación, cuidando de no introducir caracteres con acentos. Los caracteres que representan la pronunciación, están basados en la lista de fonemas utilizados para el corpus de voz en el proyecto de tesis (Miranda, 2003).

Con el archivo de palabras generadas y el archivo de universo de palabras (lista de palabras y diccionario de pronunciación, respectivamente), se generó el diccionario de pronunciación para el corpus de voz que se utiliza en este trabajo.

La instrucción para crear el nuevo diccionario es:

```
HdMan -m -w <lista_de_palabras> -n monophones1 -l dlog <nombre_nuevo_diccionario> <nombre_diccionario_de_pronunciación_global>
```

El resultado de esta instrucción es el nuevo diccionario y un archivo (monophones1) en el que se encuentran los fonemas que se utilizan en el nuevo diccionario.

Una vez determinados los datos que representan la voz, se describe a continuación la preparación de las grabaciones de voz.

Mediante la herramienta de HTK llamada HSLab, se carga en su interfaz un archivo de audio, permitiendo así observar el espectro de voz, escuchar el contenido y etiquetar segmentos de voz. Al realizar esta acción, se generó un nuevo archivo de etiquetas, cuyo contenido es el tiempo de inicio, el tiempo final y la palabra correspondiente al segmento de espectro seleccionado, es decir, se realizó una transcripción a nivel palabra.

Con los archivos de etiqueta creados, se generó un nuevo archivo, llamado también MLF (Master Label File) el cual contiene la información completa de todos los archivos de etiqueta.

Antes de continuar con la descripción del proceso de transcripción, es necesario mencionar que se llama alfabeto fonético al conjunto de símbolos que representan fonemas y que es utilizado para transcribir datos de voz, entonces, la transcripción fonética significa asociar un símbolo del alfabeto fonético con un sonido de voz (Miranda, 2003).

El proceso siguiente, fue realizar la transcripción a nivel fonético, es decir, reemplazar cada letra de cada palabra por su fonema de pronunciación.

Esta transcripción a nivel fonético se realizó con la siguiente instrucción:

```
HLEd -l '*' -d <diccionario> -i phones0.mlf  
mkphones0.led <nombre_archivo_MLF>.  
mlf
```

Donde:
phones0.mlf es el archivo generado con esta instrucción y contiene el tiempo de inicio, el tiempo final y el fonema de pronunciación que corresponde a ese segmento de tiempo.
Mkphones0.mlf es un archivo script que utiliza la herramienta.

Fig. 1. Ejemplo del archivo config

```
# Coding parameters  
TARGETKIND = MFCC_0  
TARGETRATE = 100000.0  
SAVECOMPRESSED = T  
SAVEWITHCRC = T  
WINDOWSIZE = 250000.0  
USEHAMMING = T  
PREEMCOEF = 0.97  
NUMCHANS = 26  
CEPLIFTER = 22  
NUMCEPS = 12  
ENORMALISE = F
```

Hasta este momento, la transcripción se ha realizado con las palabras que se encuentran escritas en el archivo MLF, pero hay que contemplar que en el lenguaje humano existen pausas entre palabras, es por eso que se aplica nuevamente la herramienta HLEd, cuidando de incluir en el archivo script el fonema que representa el silencio.

Por último, se parametriza las grabaciones de voz a un tipo de archivo que la herramienta HTK pueda procesar con mejor calidad y el MFCC (Mel Frequency Coefficient Cepstral) es el indicado. Esta parametrización se llevó a cabo de forma automática utilizando la herramienta HCopy, con la siguiente instrucción:

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

En la Fig.1 y en la Fig. 2 se presentan ejemplos de los archivos config y codetr.scp respectivamente.

```
../speechfiles/0/U-0.S-0.wav ../mfcc/U-0.S-0.mfc  
../speechfiles/0/U-0.S-1.wav ../mfcc/U-0.S-1.mfc  
../speechfiles/0/U-0.S-2.wav ../mfcc/U-0.S-2.mfc  
../speechfiles/0/U-0.S-3.wav ../mfcc/U-0.S-3.mfc  
../speechfiles/0/U-0.S-4.wav ../mfcc/U-0.S-4.mfc  
../speechfiles/0/U-0.S-5.wav ../mfcc/U-0.S-5.mfc  
../speechfiles/0/U-0.S-6.wav ../mfcc/U-0.S-6.mfc  
../speechfiles/0/U-0.S-7.wav ../mfcc/U-0.S-7.mfc  
../speechfiles/0/U-0.S-8.wav ../mfcc/U-0.S-8.mfc  
../speechfiles/0/U-0.S-9.wav ../mfcc/U-0.S-9.mfc  
../speechfiles/0/U-0.S-10.wav ../mfcc/U-0.S-10.mfc  
../speechfiles/0/U-0.S-11.wav ../mfcc/U-0.S-11.mfc  
../speechfiles/0/U-0.S-12.wav ../mfcc/U-0.S-12.mfc  
../speechfiles/0/U-0.S-13.wav ../mfcc/U-0.S-13.mfc
```

Fig. 2. Ejemplo del archivo codetr.scp

2) Fase de entrenamiento de los modelos fonéticos

Los algoritmos utilizados por las herramientas de entrenamiento se basan en los Modelos Ocultos de Markov.

Lo primero es crear un archivo “prototipo”, cuyo contenido es una estructura que representa a los fonemas, donde su media y su varianza sean ceros. Es necesario utilizar un archivo de configuración y un archivo que contenga la ubicación y los nombres de los archivos parametrizados.

Se creó una carpeta en la cual se añade una copia del archivo prototipo, con el objetivo de que se guarden ahí los primeros cálculos del entrenamiento.

Se utilizó la herramienta HCompV, la cual lee un conjunto de datos, calcula su media y varianza global almacenándolos en un HMM.

La instrucción es la siguiente:

```
HCompV -C <nombre_archivo_configuración> -f 0.01 -m -S <nombre_mfc> -M  
<nombre_carpeta> <nombre_archivo_prototipo>
```

Esta instrucción generó dos nuevos archivos dentro de la carpeta (la carpeta fue nombrada hmm0), donde los ceros se han sustituido por las medias y varianzas globales de la voz y servirán para calcular de manera más exacta las medias y varianzas de cada fonema.

El archivo prototipo representa la probabilidad de transición de un estado a otro y el archivo vFloors representa la densidad de de probabilidad, esta última fue hallada con una varianza global de 0.01 veces, tal como fue descrita en la instrucción. Posteriormente, se crean las definiciones de los fonemas basados en el prototipo, de tal manera que se reentrenen ahora de manera independiente. El proceso para crear las definiciones es el siguiente:

1. Se crea un nuevo archivo llamado hmmdefs dentro de la carpeta hmm0.
 - Se copia el archivo monophones0 dentro de esta carpeta (este archivo fue generado durante el proceso de creación del diccionario de pronunciación, sólo que este no incluye el fonema que representa al silencio).
 - Se renombra el archivo monophones0 a hmmdefs.
2. En cada fonema dentro del archivo hmmdefs:
 - Se coloca a cada fonema entre comillas dobles.
 - Se añade '~h' antes de cada fonema.
 - Se copia a partir de la línea 5 en adelante del archivo proto que se encuentra en la carpeta hmm0 y se pega después de cada fonema. (<BEGINHMM> a <ENDHMM>)
 - Al final del archivo se deja una línea en blanco.
3. Se crea el archivo macro:
 - Se crea un archivo llamado macro dentro de la carpeta hmm0.
 - Se copia vFloors a macros.
 - Se copia las tres primeras líneas del archivo proto a al inicio del archivo macros. (~o a <DIAGC>

Se reestimaron los cálculos, sin embargo, ahora se calculó para cada fonema, almacenando los resultados en una nueva carpeta (llamada hmm1). La herramienta HERest fue utilizada para este proceso. La instrucción para estimar (ahora a cada fonema en particular) es:
HERest -C <nombre_archivo_configuración> -I <nombre_archivo_MLF> -t 250.0 150.0 1000.0 -S <nombre_archivo_mfc> -H hmm0/macros -H hmm0/hmmdefs -M hmm1 monophones0

Se repite el mismo proceso dos veces más, creando los modelos dentro de las carpetas hmm2 y hmm3

Considerando que el lenguaje humano utiliza pausas para separar las palabras, se añadió el fonema que representa el silencio. Primeramente se copia el contenido de la carpeta hmm3 a hmm4 y usando un editor se añade el fonema "sp" de la siguiente forma:
Se copia y se pega el modelo del fonema "sil" y se renombra por "sp"
Se eliminan el estado 2 y 4 del nuevo fonema "sp", saltando el estado central.
Se cambia el parámetro de <NUMSTATES> a 3.
Se cambia el parámetro de <STATE> a 2
Se cambia el parámetro de <TRANSP> a 3

Se cambia la matriz en <TRANSP> a un arreglo de 3 x 3
Se cambia los números de la matriz como se muestra a continuación:
Se utilizó la herramienta HHed para añadir la transición extra para enlazar el estado "sp" al estado "sil" y se reentrenan los modelos fonéticos.

La instrucción es la siguiente:
HHed -H hmm4/macros -H hmm4/hmmdefs -M hmm5 sil.hed monophones1

Donde:
macros y hmmdefs son los últimos archivos generados en la reestimación.
hmm5 es una nueva carpeta que contendrá los archivos macros y hmmdefs con el nuevo estado "sp" (silencio).
sil.hed es un script cuyo contenido se muestra en la
monophones1 es el archivo generado durante la creación del diccionario de pronunciación.

Se reestimó dos veces más con la herramienta HERest los archivos generados en la carpeta hmm5, creando así los modelos dentro de hmm6 y hmm7.

Por último, se alinean los modelos de los fonemas, es decir, se toman los modelos calculados en la carpeta hmm7 y se utiliza para transformar el archivo MLF a nivel fonema usando el diccionario de pronunciación. La diferencia entre esta operación y la realizada para crear el archivo MLF, es que el reconocedor considera todas las pronunciaciones por cada palabra y produce la información con los mejores datos acústicos. La instrucción es la siguiente:

```
HVite -I '*' -o SWT -b silence -C config -a -H hmm7/macros -H  
hmm7/hmmdefs -i aligned.mlf -m -t 250.0 -y lab -I <archivo_MLF>  
-S <nombre_mfc> <diccionario_pronunciacion> monophones1
```

Se reestima dos veces más con HERest, utilizando el archivo generado por la herramienta HVite (aligned.mlf), creando los modelos en hmm8 y hmm9.

Una vez reestimados los modelos fonéticos ya alineados, ya contamos con el corpus de voz.

3) Fase de Pruebas

Esta fase se realizó con la ayuda de una interfaz gráfica y fue aplicado en niños en edad escolar. El proceso se describe más adelante.



VI. Conclusiones

En este trabajo se describió el concepto y las etapas necesarias para la creación de un corpus de voz de niños en edad escolar, que corresponden al diseño, grabación y transcripción; se describió también el proceso de creación de los modelos fonéticos que requiere el reconocedor; este proceso corresponde a las etapas de preparación de datos, entrenamiento (modelos ocultos de Markov), pruebas y análisis.

Por último se presentaron los resultados obtenidos al aplicar el sistema de reconocimiento de voz en niños con y sin problemas de lenguaje. Los resultados nos muestran que el corpus de voz reconoce satisfactoriamente las voces de niños con y sin problema de lenguaje.

El objetivo de este trabajo fue desarrollar un corpus de voz en español con niños en edad escolar y aplicarlo en una interfaz gráfica interactiva para niños con y sin problemas de lenguaje. Esto nos permitirá desarrollar aplicaciones que puedan ser útiles en la rehabilitación de niños con problemas de lenguaje.

Referencias

ávila V. y Ferrer M. (2001) Análisis Comparativo de Dos Sistemas de Reconocimiento de Voz de Habla Discreta en Personas con Alteraciones del Habla, Ponencia presentada a ISAAC 2001, Universidad de Valencia, España.

Castañeda, P. F. (1999) El Lenguaje Verbal del Niño ¿Cómo Estimular, Corregir y Ayudar para que Aprenda a Hablar Bien?, Universidad Nacional Mayor de San Marcos, Perú.

Guzmán, M. (2004) Sintetizador de Voz para la Enseñanza de la Lectura a Niños Mexicanos, Tesis Licenciatura en Ingeniería en Sistemas Computacionales, Universidad de las Américas, Puebla.

Luna M., T. (2001) Diseño e Implementación de un Sistema Tutorial basado en Tecnologías de Voz para la Enseñanza de Vocabulario, Tesis Licenciatura en Ingeniería en Sistemas Computacionales, Universidad de las Américas, Puebla.

Miranda-Palma, C. A. (2003) Sistema de Navegación Robótica por medio de Comandos Vocales con Detección Automática de Obstáculos, Tesis Maestría, ITESM, Campus Cuernavaca, Maestría en Ciencias Computacionales.

Villaseñor L. (2000) Plataforma de adquisición y análisis de datos para sistemas multimodales. Departamento de Ciencias de la Computación. Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas.

Villaseñor L., Massé A. y Pineda L. (2001) The DIME corpus, Memorias del 3er Encuentro Internacional de Ciencias de la Computación. pp 599-600, SMCC-INEGI, México.

HTK (2008), The HTK Book (for the versión 3.2.1), Cambridge University Engineering Department

RAE (2008), www.rae.es/rae/gestores/gespub000007.nsf/voTodosporId/E1F29760A46C080CC1256AD2004D606D?OpenDocument

VOXFORGE (2008), www.voxforge.org/home/dev/acousticmodels/windows/create/htkjulius/tutorial

Sobre los autores

Glendy Perera-Góngora, 28 años, nace el gusto por la investigación al trabajar en el proyecto de tesis de la licenciatura orientado hacia el reconocimiento de voz y participando paralelamente en el desarrollo del proyecto denominado Fonetix, siguiendo la misma línea de investigación, ambos proyectos orientados al apoyo a niños con discapacidad de lenguaje. Actualmente soy profesora de la Facultad de Matemáticas en la Universidad Autónoma de Yucatán (UADY), también laboro en el nivel medio superior (Conalep), y cuento con certificaciones de competencia laboral en el área de la enseñanza.

Carlos Miranda-Palma, Investigador de 36 años. Desde julio de 2000 soy profesor de la Facultad de Matemáticas – Unidad Tizimín de la Universidad Autónoma de Yucatán (UADY). He participado en proyectos del área de reconocimiento de voz desde mis estudios de maestría en el ITESM (2003) y actualmente en proyectos de investigación del cuerpo académico de Ciencias de la Computación – Unidad Tizimín. En este último hemos desarrollado proyectos de investigación con el objetivo de apoyar a personas (niños y adultos) que padecen algún tipo de discapacidad, buscando que el trabajo de investigación tenga una retribución directa hacia la sociedad. Mis principales áreas de interés son el reconocimiento de voz y la interacción humano computadora. He publicado diversos artículos en foros nacionales e internacionales. Actualmente coordinador de la Licenciatura en Ciencias de la Computación de la Unidad Tizimín y responsable del cuerpo académico de la misma.