

Creación de un sistema experto clasificador de estructuras silábicas para tareas de reconocimiento de voz mediante sílabas

M. en C. José Luis Oropeza Rodríguez, Dr. Sergio Suárez Guerra, M. en C. Ricardo Barrón Fernández,

Centro de Investigación en Computación.
Departamento de Sistemas Digitales.
Instituto Politécnico Nacional.

Av. Juan de Dios Batís esq. Miguel Otón de Mendizábal, s/n,
j_oropec2002@hotmail.com, ssuarez@cic.ipn.mx, rbarron@cic.ipn.mx.

Resumen .El presente artículo presenta los resultados obtenidos al realizar la programación de un sistema experto para realizar la división de un conjunto de palabras, frases y un archivo de texto en Español en unidades silábicas de acuerdo a las reglas que el idioma tiene para la generación de sílabas dentro de este lenguaje.

La tarea principal del sistema experto es tomar la información proveniente de palabras o frases escritas. con tal información cual se encarga de aplicar las reglas de división silábica para realizar la búsqueda de las diferentes sílabas en el medio de entrada^[1]. Dentro del motor de inferencia del sistema experto se encuentran programadas las reglas. La base de conocimiento en este caso es dinámica al igual que el mecanismo de inferencia.

El sistema fue creado con el fin de servir como material de apoyo a las aplicaciones de reconocimiento de voz que utilizan a la sílaba como unidad básica de reconocimiento^[2,3,4,5,6,7].

La finalidad es analizar el corpus de entrada y crear todo el conjunto de información de las características de tales unidades, permitiendo con ello establecer el número de elementos de éstas y con ello realizar una conexión con la base de datos del sistema para descargar en ella la información necesaria de cada uno de estos elementos, tanto de los tiempos de inicio y de etiquetado del mismo.

Los valores alcanzados al hacer uso de este sistema experto son altísimos del orden del 98 al 100% dependiendo del tipo de entrada al mismo.

Palabras clave: sílabas, Reconocimiento de voz, sistema experto, reglas de división silábica,

1. Introducción

Como parte esencial del desarrollo del presente trabajo se comenta la elaboración de un sistema experto dedicado a realizar las tareas de clasificación de sílabas en la etapa de entrenamiento del sistema. En este caso, el uso de un sistema experto es con el fin de poder hacer uso de las reglas establecidas para el Español y tratar de acoplar las necesidades de reconocimiento a tales funciones^[2,3,4].

Bajo el término de Sistemas Expertos se entiende un nuevo tipo de software que imita el comportamiento de un experto humano en la solución de un problema. Pueden almacenar conocimientos de expertos para un campo determinado y solucionar un problema mediante deducción lógica de conclusiones. En este caso se trata de facilitar la labor de un experto clasificador de sílabas, que recibe la información de un texto de entrada y extrae de él un análisis cuantitativo y cualitativo del mismo.

2. Metodología

CARACTERÍSTICAS DE LOS SISTEMAS EXPERTOS.

Los Sistemas Expertos son una expresión de los sistemas basados en el conocimiento. Con la aplicación de técnicas de Inteligencia Artificial finaliza la transición del procesamiento de datos al procesamiento de conocimientos.

Los sistemas expertos se aplican por norma general en problemas que implican un procedimiento basado en el conocimiento. Un procedimiento de solución basado en el conocimiento comprende las siguientes capacidades:

- Utilización de normas o estructuras que contengan conocimientos y experiencias de expertos especializados.
- Deducción lógica de conclusiones.
- Capaz de interpretar datos ambiguos.
- Manipulación de conocimientos afectados por valores de probabilidad.

LA FUNCIÓN DE UN SISTEMA EXPERTO

La función de un Sistema Experto es la de aportar soluciones a problemas, como si de humanos se tratara, es decir, capaz de mostrar soluciones inteligentes. Esto es posible gracias a que el sistema se crea por expertos (humanos), que intentan estructurar y formalizar conocimientos poniéndolos a disposición del sistema, para que éste pueda resolver una función dentro del ámbito del problema, de igual forma que lo hubiera hecho un experto^[8,9,10].

Poder acceder a los conocimientos adquiridos por experiencia es lo más difícil, ya que los expertos, al igual que otras personas, apenas los reconocen como tales. Son buscados con mucho esfuerzo y cuidado siendo descubiertos de uno en uno, poco a poco.

COMPONENTES DE UN SISTEMA EXPERTO

Una característica decisiva de los Sistemas Expertos es la separación entre conocimiento (reglas, hechos) por un lado y su procesamiento por el otro. A ello se añade una interfaz de usuario y un componente explicativo.

- La Base de Conocimientos de un Sistema Experto contiene el conocimiento de los hechos y de las experiencias de los expertos en un dominio determinado.
- El Mecanismo de Inferencia de un Sistema Experto puede simular la estrategia de solución de un experto.
- El Componente Explicativo explica al usuario la estrategia de solución encontrada y el por qué de las decisiones tomadas.
- La Interfaz de Usuario sirve para que éste pueda realizar una consulta en un lenguaje lo más natural posible.
- El Componente de Adquisición ofrece ayuda a la estructuración e implementación del conocimiento en la base de conocimientos.

LA BASE DE CONOCIMIENTOS

La Base de conocimientos contiene todos los hechos, las reglas y los procedimientos del dominio de aplicación que son importantes para la solución del problema^[8,9,10].

La base de conocimientos dispone de reglas. Estas reglas se representan en forma de:
Si premisas Entonces Conclusión y/o Acción.

En la zona de las premisas se solicitan vinculaciones lógicas referentes a las cualidades de los objetos.

En la zona de la conclusión se añaden nuevos hechos y cualidades a la base de conocimientos y/o se ejecutan acciones. Esto se define a menudo como programación orientada a reglas.

En nuestro caso la base de conocimientos se encuentra constituida por todas las reglas de clasificación de sílabas del lenguaje Español que se analizan a continuación:

LAS REGLAS DE LA SÍLABA.

Palabras monosilábicas:

Son las que están formadas por una sílaba: luz, mar.

Palabras bisilábicas:

Son las que están formadas por dos sílabas: silla, mesa.

Palabras trisilábicas:

Son las que están formadas por tres sílabas: ventana, cabeza.

Palabras polisilábicas:

Son las que están formadas por cuatro o más sílabas: Argentina, Polideportivo.

En ^[12] se menciona que en el Español existen 29 letras, las cuales están clasificadas de acuerdo a su pronunciación en dos grupos: vocales y consonantes. En la figura 1 se muestra un esquema de la clasificación de las letras para una mejor comprensión. El grupo de las vocales está formado por cinco letras, su pronunciación no dificulta la salida del aire. La boca actúa como una caja de resonancia abierta en menor o mayor grado y de acuerdo a esto, las vocales se clasifican en abiertas, semiabiertas y cerradas^{[2][3]}.

El otro grupo de letras, las consonantes, está formado por veinticuatro letras, de las cuales tres son consonantes compuestas, llamadas así, por ser letras simples en su pronunciación y dobles en su escritura. Las letras restantes son llamadas consonantes simples, por ser simples en su pronunciación y en su escritura.

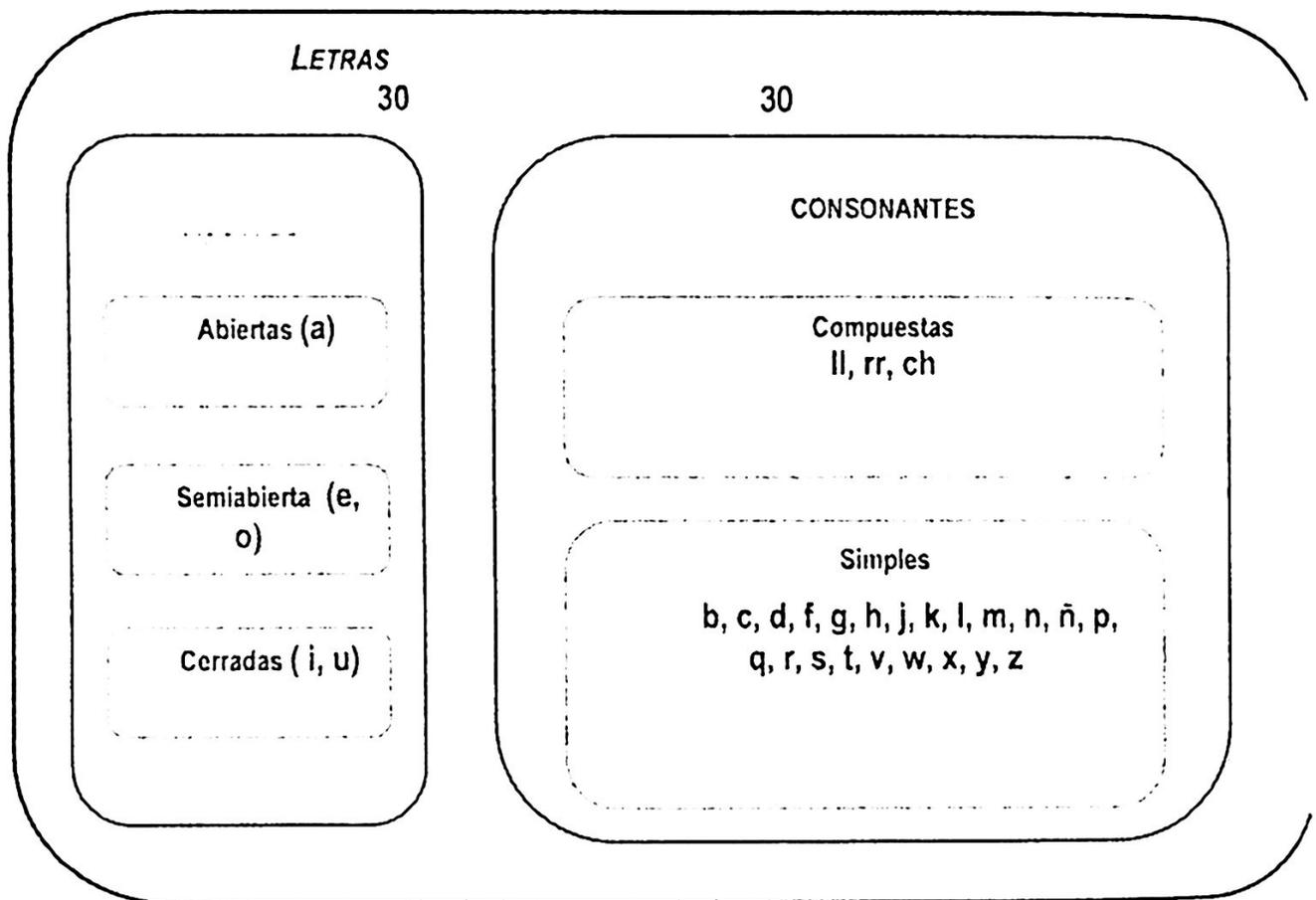


Fig. 1 Esquema de la clasificación de las letras.

Las sílabas

En el idioma español existen once reglas, las cuales determinan la separación de las sílabas de una palabra. Estas reglas son listadas a continuación mostrando enseguida excepciones a la misma.

REGLA 1

En las sílabas, por lo menos, siempre tiene que haber una vocal. Sin vocal no hay sílaba.

REGLA 2

Cada elemento del grupo de consonantes inseparables, mostrado en la figura 3.6, no puede ser separado al dividir una palabra en sílabas.

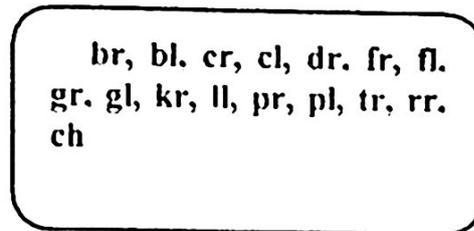


Fig. 3.6. Grupo de consonantes inseparables.

REGLA 3

Cuando una consonante se encuentra entre dos vocales, se une a la segunda vocal.

REGLA 4

Cuando hay dos consonantes entre dos vocales, cada vocal se une a una consonante.

Excepción: Esto no ocurre en el grupo de consonantes inseparables (regla 2).

REGLA 5

Si son tres las consonantes colocadas entre dos vocales, las dos primeras consonantes se asociarán con la primera vocal y la tercera consonante con la segunda vocal.

Excepción. Esta regla no se cumple cuando la segunda y tercera consonante, forman parte del grupo de consonantes inseparables.

REGLA 6

Las palabras que contienen una h precedida o seguida de otra consonante, se dividen separando ambas letras.

REGLA 7

El diptongo es la unión inseparable de dos vocales. Se pueden presentar tres tipos de diptongos posibles:

- Una vocal abierta + Una vocal cerrada.
- Una vocal cerrada + Una vocal abierta.
- Una vocal cerrada + Una vocal cerrada.

Son diptongos sólo las siguientes parejas de vocales: ai, au, ei, eu, io, ou, ia, ua, ie, ue, oi, uo, ui, iu, ay, ey, oy .

La unión de dos vocales abiertas o semiabiertas no forma diptongo, es decir, deben separarse en la segmentación silábica. Pueden quedar solas o unidas a una consonante.

REGLA 8

La h entre dos vocales, no destruye un diptongo. *Ejemplo:*

Palabra	sílabas
ahuyentar	ahu + $\bar{y}\bar{e}\bar{n}$ + tar ↑↑↑

REGLA 9

La acentuación sobre la vocal cerrada de un diptongo provoca su destrucción.

Palabra	sílabas
María	Ma + $\bar{r}\bar{i}$ + a ↑↑

REGLA 10

La unión de tres vocales forma un triptongo. La única disposición posible para la formación de triptongos es la siguiente:

Vocal cerrada + (vocal abierta | vocal semiabierta) + vocal cerrada

Sólo las siguientes combinaciones de vocales, forman un triptongo: iai,iei, uai, uei, uau, iau, uay, uey.

3. El mecanismo de inferencia.

El mecanismo de inferencia es la unidad lógica con la que se extraen conclusiones de la base de conocimientos, según un método fijo de solución de problemas que está configurado imitando el procedimiento humano de los expertos para solucionar problemas. Una conclusión se produce mediante aplicación de las reglas sobre los hechos presentes^[8,9,10].

Ejemplo.

Una Regla es: Si p y q entonces r

p y q son justo aquellos hechos que se mencionan en la cláusula "si" de la regla, es decir, las condiciones para la aplicabilidad de la regla. Aplicar la regla es: deducir de los hechos p y q el hecho r.

En un Sistema Experto existirá un hecho sólo cuando esté contenido en la base de conocimientos.

Los hechos que constan en la cláusula "si" se llaman *premisas*, y el contenido en la cláusula "entonces", se llama *conclusión*. Cuando se aplica una regla sobre algunos hechos cualesquiera se dice que se *dispara*. El disparo de una regla provoca la inserción del nuevo hecho en la base de conocimientos.

Las funciones del mecanismo de inferencia son:

- Determinación de las acciones que tendrán lugar, el orden en que lo harán y cómo lo harán entre las diferentes partes del Sistema Experto.
- Determinar cómo y cuándo se procesarán las reglas, y dado el caso también la elección de qué reglas deberán procesarse.
- Control del diálogo con el usuario.

La decisión sobre los mecanismos de procesamiento de reglas, es decir, qué estrategias de búsqueda se implementarán, es de vital importancia para la efectividad del sistema en su conjunto.

4. Resultados

La implantación del sistema experto para el presente trabajo tiene como entrada el conjunto de frases o palabras que conforman un vocabulario determinado a reconocer, tras la aplicación de las reglas pertinentes, se procede a corroborar cada uno de los elementos de entrada, con lo cual se puede lograr establecer los inicios y finales de una sílaba y la otra.

Lo anterior cumple con el objetivo de que en el reconocimiento se extraigan la cantidad y tipos de sílabas que conforman el corpus y asimismo provee la cantidad de bloques que serán usados en la etapa de segmentación y etiquetado silábico de las señales de voz. Todo lo anterior, permite encontrar e indicar un parámetro de referencia en los siguientes puntos:

Cuando se utiliza en el entrenamiento, se procede a concordar el número de sílabas extraídas del experto con el número obtenido de las estructuras energéticas que se analizan para realizar tal segmentación y clasificación. En caso de no coincidir se desecha la muestra analizada y se retoma una nueva.

Cuando se lleva a cabo el reconocimiento del habla discontinua, se hace uso de tal segmentación, para extraer cada uno de los componentes de las frases o de las sílabas para ser analizados de forma independiente por cada uno de los algoritmos de reconocimiento y establecer por tanto la decisión de cual será la decisión a tomar por el sistema de reconocimiento.

IMPLANTACIÓN DEL SISTEMA EXPERTO.

Básicamente y debido a que la finalidad es crear el sistema experto que se acople a las necesidades del tipo de reconocimiento que estamos tratando de realizar, se procedió a realizar el diseño y la programación del mismo usando lenguaje C para ello. Las principales características de tal programa se describen a continuación^[8,9,10]:

- Posee una toma de decisión de tipo cualitativo y cuantitativo del tipo de sílabas que comprenden la entrada que se le está otorgando.
- La entrada al sistema se produce a través del software diseñado para las aplicaciones referentes a los fines de investigación precisos.
- En caso de ser frases, el sistema se encarga de realizar su separación en palabras para después separarlas por sílabas.
- La base de conocimiento es usada para poder realizar tal tarea de segmentación, se tiene como referente las reglas del Español para la obtención de sílabas. EL uso de las sentencias if proposición then, son el estándar usado en esta capa del sistema.
- La base de programación está realizada en estructuras dinámicas de datos del tipo lista enlazada. Las cuales nos permiten ir accediendo a cada uno de los elementos de la palabra para posteriormente realizar la segmentación adecuado de la palabra según sea necesario.

A continuación se muestran dos segmentos del código del sistema experto en cuestión:

```
void CDivide_elemento::inicializa (void)
{
  posiciones_corte[0]=elemento;
  elemento++;
  for (i=0;i<20;i++) reglas[i]=false;
  cons_insepa="brblcrldrfrflgrglkrllprpltrrrchtl";
  vocal_diptongo="aiaueieuiouiauaieueoiuuiuyeyoy";
  vocal_abierta="íaíoíeíuúaúeúoúiooaeaoeaoeooaeeee";
  vocales="aeiouíúáéó";
}
```

Este método se encarga de inicializar los elementos necesarios para analizar dentro de una palabra la existencia de elementos tales como: las consonantes inseparables, las vocales diptongo, las vocales abiertas, las vocales, etcétera.

*/*aquí se comienza a analizar la regla 2, después de ella se identificará si existen
consonantes inseparables en la cadena de texto que se está analizando*/*

```
r=inicio
do
{
  ptr1=*r->fonema;
  if (ptr1=="C")
  {
    p2=r->next;
    ptr2=*p2->fonema;
    if(ptr1==ptr2)
    {
      for (i=0;i<int(strlen(cons_insepa));i+=2)
      {
        if(cons_insepa[i]==palabra[contador]
cons_insepa[i+1]==palabra[contador+1])
        {
          conta_conso_insepa[cuenta_insepa]=contador;
          cuenta_insepa++;
          reglas[1]=true;
          if(verifica_existencia(posiciones_corte,elemento,contador))
          {
            posiciones_corte[elemento]=contador;
            elemento++;
          }
        }
      }
    }
  }
  r=r->next;
  contador+=1;
}while (r->next!=NULL);
```

Las sentencias anteriores permiten observar la forma en la cual se trabaja sobre el conjunto determinado de elementos de entrada para poder realizar las tomas de decisiones adecuadas, el uso de banderas que representan las reglas que se activan conforman los resultados de la aplicación de los mecanismos de inferencia que internamente se han aplicado. Cuando estas reglas satisfechas se contrarrestan, se procede a analizarlas de forma posterior en otra etapa posterior del mismo sistema.

Los resultados obtenidos en la división silábica al hacer uso de este sistema en los diferentes corpus usados a lo largo del presente trabajo se resumen en la siguiente tabla, manifestándose en la misma su porcentaje de efectividad alcanzado.

TIPO DE CORPUS	PORCENTAJE DE EFECTIVIDAD
Silabas independientes	100%
Dígitos	100%
Corpus de palabras aisladas	100%
Corpus de Latino40	100%
Corpus de frases simples	100%

Tabla 5.1. porcentaje de efectividad en la aplicación del sistema experto a diferentes corpus de voces.

5. Discusión

La inclusión de un sistema experto a un sistema de reconocimiento de voz no es nuevo, técnicas empleadas con anterioridad de la IA (Inteligencia Artificial) demuestran este hecho, lo importante en estos momentos es la oportunidad de intercalar las reglas de las sílabas a los sistemas basados en los mismos.

Los resultados obtenidos en el presente reporte servirán para poder analizarlos e incrustarlos en tareas de reconocimiento.

6. Conclusiones

La incrustación de los sistemas expertos para las tareas de reconocimiento de voz por computadora, representa una alternativa para conseguir obtener mejores rendimientos de su uso. En este trabajo hicimos uso de las características y beneficios del sistema experto para clasificar un conjunto de elementos de entrada (frases, palabras, etc.) en unidades silábicas.

La tarea principal fue programar al experto con la base de conocimiento de las reglas de las sílabas y sobre la base de este parámetro, realizar la clasificación adecuada. Como se observa en los resultados, el experto clasificador alcanzó altos rendimientos en cuanto a identificar de forma adecuada el número de sílabas total del elemento de entrada y además asignarlas a cada una de las estructuras silábicas correspondientes.

Dado lo anterior se tiene la posibilidad de encontrar todos estos parámetros una vez que se ha alimentado al sistema de forma adecuada.

Referencias

- [1]Speech Recognition Using Syllable-Like Units, Zhihong Hu, Johan Schalkwyk, Etienne Barnard, Ronald Cole. Center for Spoken Language Understanding. Oregon Graduate Institute of Science and Technology. 1998.
- [2]Análisis del Español Mexicano, para la construcción de un sistema de reconocimiento de dicho lenguaje, Ben Serridge, GRUPO TLATOA, UDLA, Puebla México 1998.
- [3]Syllable Structure Constraints, A C/D Model Perspective, Osamu Fujimura. 1998.

- [4] Sobre el uso de la sílaba como unidad de síntesis en el Español. Feal L. Pinto, Informe técnico, Departamento de Informática, Universidad de Valladolid, 2000.
- [5] Continuous speech recognition using automatically segmented data at syllable units. V. Kamakshi Prasad, T. Nagarajan and Hema A. Murphy. Department of Computer Science and Engineering, Indian Institute of Technology, Madras, Chennai, ICSP, 2002.
- [6] Integrating Syllable Boundary Information Into Speech Recognition. Su-Lin Wu, Michael .
- [7] Pruebas y validación de un sistema de reconocimiento del habla basado en sílabas con un vocabulario pequeño. Sergio Suárez Guerra, Karen Suso, Mariana del Villar, Congreso Internacional de Computación CIC2003. México, D.F.
- [8] Jesus Savage Carmona, A Hybrid Systems with Symbolic AI and Statical Methods for Speech Recognition, University of Washington, 1995.
- [9] Stuart Russell, Peter Norvig, Inteligencia Artificial un enfoque moderno, Prentice Hall, 1996.
- [10] Joseph Giarratano y Gary Riley, International Thompson Editores, Sistemas expertos, principios y programación 2001.

José Luis Oropeza Rodríguez; Centro de Investigación en Computación (CIC); IPN; Av. Juan de Dios Batis esq. Miguel Otón de Mendizábal, s/n, México; email: j_oro2002@hotmail.com; Tel: 57296000 ext. 56519

Sergio Suárez Guerra; Centro de Investigación en Computación (CIC); IPN; Av. Juan de Dios Batis esq. Miguel Otón de Mendizábal, s/n, México; email: ssuarez@cic.ipn.mx; Tel: 57296000 ext. 56519

Ricardo Barrón Fernández; Centro de Investigación en Computación (CIC); IPN; Av. Juan de Dios Batis esq. Miguel Otón de Mendizábal, s/n, México; email: rbarron@cic.ipn.mx; Tel: 57296000 ext. 56519

Biografía corta de cada autor

José Luis Oropeza Rodríguez, nació el 30 de Noviembre de 1972. Graduado de Ingeniería en Telecomunicaciones y Electrónica en la ESIME UC, en 1994. Graduado de Master en Ingeniería de Cómputo en el CIC en 2000. Se desempeña como alumno de doctorado en el CIC, IPN desde 2000. Su área de interés es el reconocimiento de voz.

Dr. Sergio Suárez Guerra. Graduado de Ingeniero Electricista en el Instituto Politécnico José Antonio Echeverría en La Habana, Cuba en 1972, Doctor en Ciencias en el Centro de Cálculo de la Academia de Ciencias de la URSS en 1979. Profesor Investigador del Instituto Politécnico Nacional desde 1998. Su área de interés: Procesamiento digital de señales, en especial Voz.

M. en C. Ricardo Barrón Fernández. Graduado de Matemático en la Facultad de Ciencias de La Universidad Nacional Autónoma de México en 1985, Maestro en Ciencias en el Centro de Investigación en Computación del Instituto Politécnico Nacional (IPN) en 1998. Profesor Investigador del IPN desde 1986. Su área de interés: Procesamiento digital de señales, imágenes y voz.

71