

Hoy en día los sistemas de reconocimiento de voz son aplicados en diferentes áreas como telefonía, sistemas de seguridad, interacción con computadoras, etc. Sin embargo se requiere que estos sistemas tengan un buen desempeño (robustos) y una interacción más natural. Para esto se han desarrollado técnicas que permiten a los sistemas entablar comunicación de una manera fácil con los humanos a través de la voz. Algunas de las técnicas de reconocimiento están basadas en redes neuronales y en modelos ocultos de Markov.

1.1 Los sistemas de reconocimiento automáticos de voz

Los métodos de reconocimiento automático de voz han sido investigados por muchos años, enfocándose principalmente en aplicaciones dedicadas a dar información por teléfono. En nuestros días se ha intensificado dicha investigación y los logros que se han obtenido aún son muy limitados en su uso. Por esta razón se han estudiado métodos que aumenten la capacidad de reconocimiento en los sistemas para que sean robustos y tengan una mejor adaptación al medio.

1.1.1 Definición

Un sistema de reconocimiento de voz es el proceso de extraer automáticamente y determinar la información lingüística dado por un archivo de voz usando computadoras o circuitos electrónicos [FURUI89]. El objetivo de dichos reconocedores es convertir las palabras habladas a palabras escritas.

En un principio en los sistemas de reconocimiento de voz típicos, la voz de entrada era comparada con modelos de referencia guardados en fonemas o palabras, y el modelo que más se parezca es seleccionado como un fonema o palabra candidato de la entrada de voz.

Algunas de las dificultades que se presentan en el reconocimiento de voz son el modelado de fenómenos coarticulatorios y la reducción de algunos problemas como ruido y disfluencias. Una de las principales es la coarticulación, que es el período de transición de un fonema a otro, es decir, los órganos articulatorios se mueven al lugar de articulación del

fonema siguiente, antes de completar la articulación del fonema presente. A la identidad de estos fonemas se les denomina contexto, por lo que se considera diferente un fonema con un contexto que el mismo con otro contexto, para modelar este tipo coarticulatorio se usan diferentes tipos de unidades:

- Fonemas divididos en una parte, que son aquellos en los que sólo se toman en cuenta uno de los contextos y se dividen en:
 - a. dependiente-izquierdo: que dependen únicamente del contexto del lado izquierdo, y
 - b. dependiente-derecho: que dependen únicamente del contexto del lado derecho.

- Los fonemas divididos en 2 partes son aquellos que dependen de ambos contextos, el derecho y el izquierdo.

1.2 Clasificación del reconocimiento de voz

El reconocimiento de voz puede ser clasificado en reconocimiento de propósito específico o de propósito general. El enfoque que se da en esta tesis está orientada hacia el segundo, ya que se hará un reconocedor de propósito general el cual reconocerá palabras de cualquier dominio.

1.3 Los componentes de un reconocedor automático de voz.

Para construir un reconocedor de voz es necesario obtener una base de datos de voz con un dominio específico. Esto es, se necesita un conjunto de archivos de voz que posteriormente se digitalizan y se extrae el conjunto de características para introducirlas a un clasificador el cual se encarga de obtener ciertas probabilidades para cada conjunto de características que se introduzcan.

Al obtener las probabilidades y con una estructura definida que nos dé las pronunciaciones posibles que deseamos reconocer, se realiza un algoritmo de búsqueda para encontrar la secuencia más probable, y así lograr un nivel de reconocimiento satisfactorio. Todo este proceso se puede ver en la *figura 1.1*.

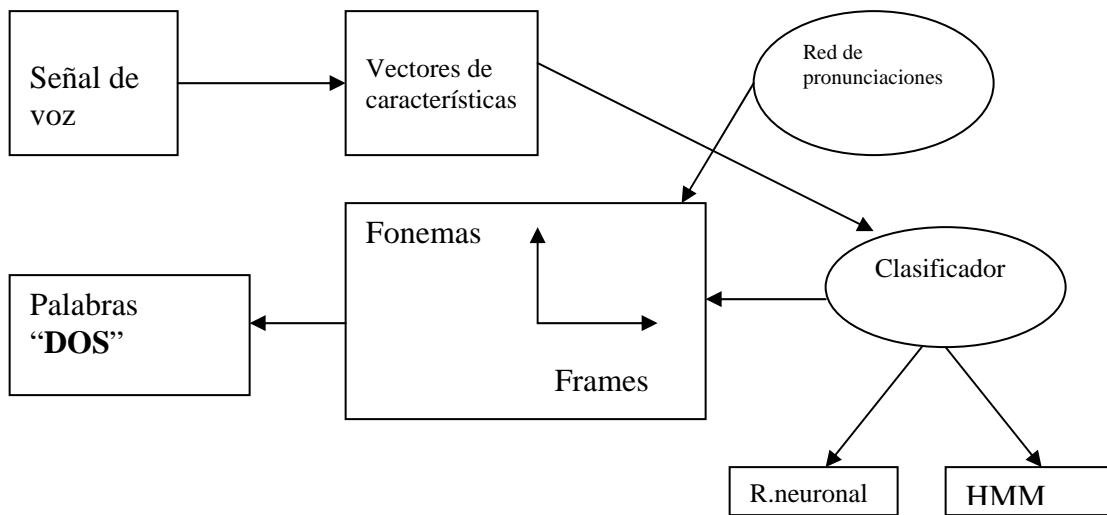


Figura 1.1 Estructura general de un reconocedor automático de voz

En esta tesis se enfocará a el problema de clasificación. Existen dos enfoques para el reconocimiento de voz muy utilizados, los cuales son redes neuronales y modelos ocultos de Markov. Se analizarán ambos enfoques y se estudiará el desempeño obtenido de cada uno de ellos. Para esto tomaremos en cuenta características importantes como el tiempo total de las fases de entrenamiento, que tipo de software y hardware fueron ocupados para cada experimento y el tipo de errores comunes que se presentaron en los desarrollos de los mismo. Ahora daremos una pequeña introducción sobre redes neuronales y modelos ocultos de Markov. En los siguientes capítulos se hablará detalladamente de cada uno de los enfoques mencionados.

1.4 Redes neuronales

Básicamente podemos decir que entre las topologías más conocidas sobre redes neuronales tenemos:

- Perceptrones multicapa o de una sola capa
- Redes recurrentes y de Hopfield
- Redes de Kohonen

1.4.1 Algunas topologías

En el perceptrón, las salidas de uno o más elementos computacionales en una sola capa forman las entradas hacia un nuevo conjunto de elementos computacionales simples de la siguiente capa.

Hopfield propuso varios modelos de redes recurrentes. En este tipo de redes, la salida de cada neurón se calcula y se retro-alimenta como entrada, calculándose otra vez, hasta que se llega a un punto de estabilidad, pero se pueden dar casos de que nunca se tenga un punto estable.

La tercera topología es la de Kohonen, o mapeo de características que se organizan a sí mismas, la cual es un procedimiento de reagrupación para proveer un libro de códigos de patrones estables en el espacio de entrada que caracteriza una entrada arbitraria a un vector, por un pequeño número de clusters representativos.

Características

Las características más importantes de una red son las siguientes:

- Número y tipo de entradas
- Conectividad de la red
- Opción de compensación
- Opción de no linealidad

Ventajas

Ventajas de las redes neuronales:

- Implementación rápida de un grado masivo de computación paralela
- Algunos modelos son robustos o tolerantes al fallo
- Los pesos de conexión de la red no necesitan ser restringidos
- Debido a la no linealidad dentro de cada elemento computacional, una red neuronal puede ser aproximada a cualquier no linealidad o a un sistema dinámico no lineal.

En los siguientes capítulos se detallarán aspectos importantes sobre redes neuronales para la realización de esta tesis.

1.5 Modelos Ocultos de Markov

Los modelos ocultos de Markov es una técnica probabilística para el estudio de series en el tiempo. La teoría permite modelar muchas de las distribuciones de probabilidad clásicas.

Los elementos en las series pueden ser contables o distribuidos continuamente; estos pueden ser escalares o vectores. Dichos modelos utilizan métodos estocásticos; una serie en el tiempo es generada y analizada por un modelo paramétrico de probabilidad. Es paramétrico en el sentido de que es completamente descrito por una lista finita de números reales. Un modelo de Markov tiene dos componentes muy importantes: una cadena de Markov de estados finita y un conjunto finito de salidas de distribuciones de probabilidad. Si el modelo es visto generativamente, la cadena de Markov sintetiza una secuencia de estados (llamada ruta) y la salida de las distribuciones, entonces vuelve a esta ruta una serie o series en el tiempo. Y si es vista analíticamente, una serie en el tiempo observada nos da evidencia de la ruta oculta y los parámetros de generación del modelo.

Una vez establecidos algunos de los conceptos más importantes de esta tesis, en el siguiente capítulo hablaremos sobre los enfoques de redes neuronales y modelos ocultos de Markov que son la base de los clasificadores en uno y otro reconocedor.