CAPÍTULO 2.

RECONOCIMIENTO DE VOZ y VXML

2.1 Reconocimiento de voz

Como lo menciona H. Meza (1999) en su tesis: “El habla constituye un canal de comunicación entre los humanos, físicamente se forma de la presión del aire emitida por el sistema articulatorio, para su estudio es necesario transformar la voz en una señal eléctrica mediante un micrófono y someterla posteriormente a diferentes procesos digitales”.

Un reconocedor de voz es una herramienta capaz de traducir una señal de voz a texto. Un sistema basado en voz generalmente busca entender el lenguaje verbal con el objetivo de poder recibir tareas y datos del usuario que habla, es decir, es el establecimiento de una interfaz Hombre-Máquina más natural mediante la voz humana.

En la actualidad se han desarrollado algunos sistemas que pueden llegar a reconocer vocabularios limitados en voz de ciertas personas, sin embargo es muy complejo que un sistema pueda contener una gran variedad de formas de pronunciación, acentos y combinaciones de peticiones.

Las áreas de aplicación de los sistemas de reconocimiento son diversas: discapacitados, servicios de información automática, sistemas de seguridad y la interacción con computadoras, entre otras, son algunas de estas áreas.

2.2 Arquitectura de un sistema de Reconocimiento de Voz.

Como ya se había mencionado antes los reconocedores de voz traducen una señal de voz a texto, este proceso se divide en tres áreas:

- Preprocesamiento: Convertir una entrada de voz o señal análoga a una señal digital que pueda ser procesada por el reconocedor de voz.
- Reconocimiento: Traducir la señal a texto, es decir, es la identificación del habla.
- Comunicación: Trasmitir la señal reconocida a una aplicación.

Figura 2.1 Componentes de aplicación de voz [Kirschning, 2006]

El funcionamiento de un reconocedor de voz se realiza de la siguiente forma:

1.- Capturar la señal de voz y convertirla en una señal digital.

2.- Sacar de la señal digitalizada una serie de características esenciales.

3.- Usando la serie de características y el clasificador, se saca a continuación una serie de probabilidades.

4.- Se realiza una búsqueda mediante las probabilidades y una estructura con las posibles pronunciaciones se muestra el resultado del reconocimiento de la señal de voz entrante.

2.3 Elementos para el Funcionamiento del Reconocedor de Voz.

Un reconocedor es relativamente sencillo si sólo tiene que reconocer palabras aisladas, sin embargo es más complejo si debe reconocer las palabras de una frase, pero introduciendo un pausa entre cada palabra, el sistema más complicado es aquel que debe de funcionar reconociendo habla continua o forma natural del habla. [Poza, 1991]

Para que cualquier tipo de reconocedor pueda funcionar necesita de los siguientes elementos básicos:

- Vocabulario: Es el número de palabras diferentes que debe de reconocer el sistema. Mientras más grande el número de palabras diferentes mas difícil es el reconocedor, debido a que, con mayor probabilidad puede que aparezcan palabras parecidas entre sí.
- Gramática: Es el conjunto de reglas que limita el número de combinaciones permitidas de las palabras del vocabulario. Una gramática ayuda a mejorar la tasa de reconocimiento a través de la eliminación de ambigüedades, además, de aumentar la rapidez y precisión del proceso de reconocimiento al limitar el número de palabras en una determinada fase del reconocimiento, es decir, el diseño adecuado de las reglas y la interfaz optimizan el reconocimiento. Ejemplo si una aplicación debe de reconocer un número de teléfono, la gramática de este dice que el vocabulario son los números del 1 al 10, y debe de

reconocer un conjunto de 7 dígitos, de manera que si el sistema reconoce mas o menos, quiere decir que existe algún error. [Poza, 1991]

- Idioma: Indica a el reconocedor bajo que idioma se va a estar trabajando, es necesario, ya que, en cada idioma cambia la forma de pronunciación y significado de las palabras, y por lo tanto, influye en la construcción de la gramática y vocabulario del reconocedor.

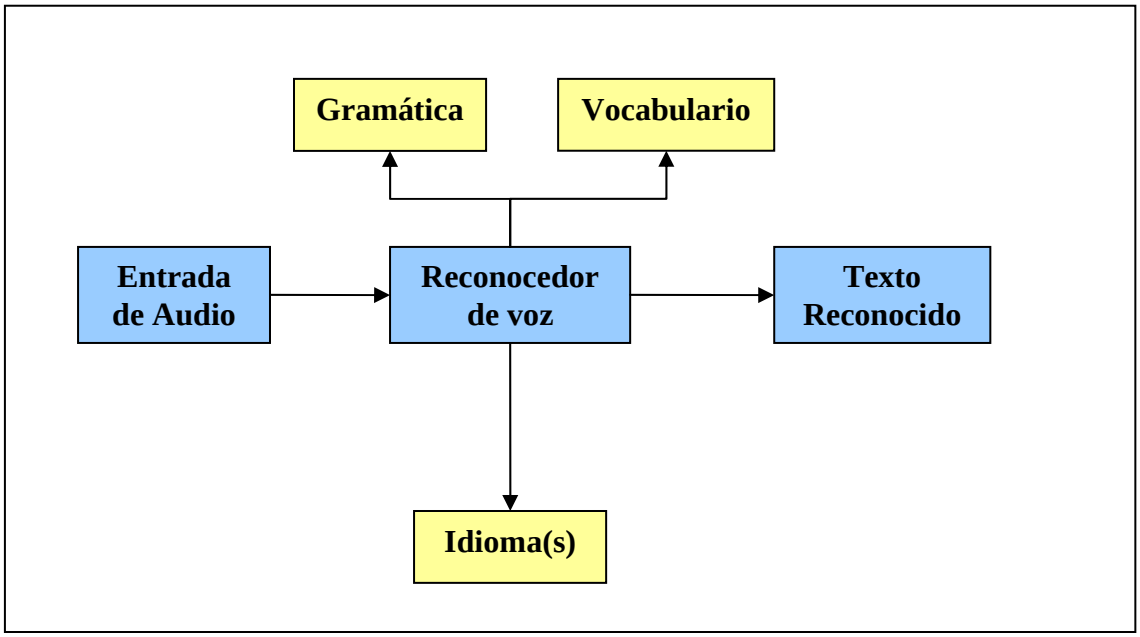

Figura 2.2 Diagrama General del Reconocedor de voz

2.4 Dificultades

Los sistemas de reconocimiento de voz nos brindan muchas ventajas, como la introducción de datos a una base de datos o bien a un procesador de texto de forma rápida y como consecuencia de esto, se evita el uso de dispositivos de entrada como el ratón y el teclado.

Sin embargo existen muchas dificultades en el proceso de reconocimiento de voz una de estas dificultades es el ruido que se capta junto con la señal de voz debido a que fuentes externas como: televisiones, radio, pláticas de otra gente y otros ruidos ajenos a la señal de voz; se encuentran cerca del micrófono encargado de trasmitir la señal de voz al reconocedor.

Otra de las dificultades es el sonido del habla emitido por cada usuario, es decir, cada persona tiene distintas formas de hablar, de hacer sonar los distintos acentos, de velocidad del habla, de respiración y de dicción, además, de la gran variedad dialéctica de hablar el idioma.

2.5 Aplicaciones

Los sistemas de reconocimiento de voz tienen como objetivo principal implementar interfaces dirigidas a las necesidades de los usuarios a través de su voz. Algunas de las aplicaciones de estos reconocedores son las basadas en el teléfono, de dictado automático y control de robots por mencionar algunas. Una de las aplicaciones de estos reconocedores más usadas y de más éxito es el teléfono, debido a su bajo costo, fácil implementación y disponibilidad.

Algunas de aplicaciones que utilizan este tipo de reconocedores por teléfono son:

- Asistencia
- Servicios Financieros
- Llamadas por cobrar.
- Información.

Un ejemplo de este tipo de aplicación se encuentra en la Universidad de las Américas Puebla, con un sistema de conmutador automático (CONMAT), sistema que

tiene la capacidad de dirigirte hacia cualquier departamento dentro de la Universidad que el usuario necesite.

2.6 Voice Extensible Markup Language (VXML).

VXML es un lenguaje de marcado basado en XML para crear aplicaciones distribuidas que permite emplear síntesis de voz, audio digitalizado, reconocimiento del habla, marcación de teclado, registro de entradas telefónicas, combinación de conversaciones y entradas DTMF (Dual Tone Multi-Frecuency). El principal objetivo de VXML es ofrecer ventajas del desarrollo basado en Web y aplicaciones de voz interactivas. [VXML, 2004].

VoiceXML proporciona un entorno abierto con una descripción de diálogos y formato de gramáticas estándar, que se encarga de convertir el habla en texto con la ayuda de las gramática de reconocimiento del habla (SGRS). [VXML, 2004].

Mientras que HTML es utilizado para la creación de interfaces graficas para que el usuario pueda ingresar y recibir información, VXML genera interfaces orales, es decir, a diferencia de HTML el usuario no ve la información la escucha, esta entrada de audio está controlada por un reconocedor de voz integrado con el navegador de VXML, es por eso que el usuario no necesita mas que un teléfono para poder acceder a este tipo de aplicaciones. VoiceXML necesita de un navegador al igual que HTML para poder reconocer y procesar las etiquetas del lenguaje, a través de un intérprete.

Algunas aplicaciones en las que se puede utilizar VXML son: sistemas de información, comercio electrónico y servicios telefónicos. Es por esto que las aplicaciones construidas con VoiceXML en combinación con otras tecnologías para Internet están teniendo un gran crecimiento en Internet. [VXML, 2004].