

Capítulo I. Introducción

1.1 Descripción del trabajo

El ser humano, como todos sabemos tiene la necesidad de comunicarse, de ser escuchado y sobretodo interactuar con los demás seres vivos que lo rodean. Así mismo de compartir sentimientos, formas de expresarse, y sentimientos expresados por medio del habla.

La comunicación es la transmisión de información de un lugar a otro, mediante un proceso en el cual se incluye un mensaje de un emisor a un receptor por medio de un canal.

La comunicación, tiene diferentes formas de ser procesada y de ser transmitida ya que llega a nosotros a través de señales acústicas o visuales, audio, impresa, sonidos, palabras, dibujos, signos etc. Esto se realiza con el fin de que la persona que este recibiendo el mensaje sea capaz de procesarlo y entenderlo, en caso contrario la persona no podrá captar lo que en el mensaje se transmite.

Desde nuestros antepasados, la necesidad de comunicación es básica para poder realizar actividades de sobre vivencia. El poder comunicarse y saber como transmitir los sentimientos, el saber escuchar y procesar la información requerida o transmitida, forma parte del lenguaje y de la comunicación oral.

Es importante mencionar las ventajas que nos ofrece la comunicación oral frente al empleo de los métodos mas tradicionales, dentro de los cuales esta el de ser mas reconfortable con los usuarios, esto es, el modo de comunicación mas natural entre los seres humanos, además de que posibilita la interacción con otras tareas de forma simultanea a dicha comunicación. Permite el acceso a la información en poco tiempo, esto hace que sea de considerar que la voz puede interferir de diferentes maneras en otros sistemas de comunicación, y puede ser fácilmente interceptable.

Dentro de la comunicación, encontramos también que es necesario establecer con la maquina una buena interacción, para esto es necesario que exista una parte que sea capaz de recopilar y descomponer los mensajes de voz u orales que sean enviados a la máquina, y por otro lado otro modulo que codifique y formule mensajes que sean entendibles por los seres humanos. Con el paso del tiempo han ido apareciendo diversas técnicas para el reconocimiento (descodificación) y de síntesis (codificación) de voz, que pretenden realizar una buena comunicación bidireccional, intentando tener una alta calidad y flexibilidad.

Es necesario nombrar los sistemas de Conversión de Texto a Voz, los cuales realizan una traducción de algún texto escrito a una representación lingüística de un mensaje, por medio de lo q se conoce como proceso lingüístico prosódico, y después de realizar este proceso pasa a representación de onda acústica, la cual es llamada como proceso acústico, con estas dos partes se da lo que se conoce como síntesis de voz.

Dentro de las características que debemos de tomar en cuenta para una buena síntesis, mencionaremos las siguientes:

- Excelente calidad de audio, en lo que se refiere a naturalidad
- Flexibilidad para que pueda emitir cualquier mensaje que le sea requerido
- Procesamiento rápido de las formas de onda.

La implementación de los sintetizadores de voz en cuanto aplicaciones reales, se van dando con mayor frecuencia, gracias a los avances que se han dado en la segunda mitad de este siglo, y no solo avances dentro de la rama de la ingeniería, si no también dentro de los conocimientos lingüísticos, estos a su vez, nos exige un alto nivel de conocimiento fonético, ya que sin esto no podríamos alcanzar un alto grado que nos permita una buena adecuación con las aplicaciones.

Dentro de las aplicaciones desarrolladas podríamos hacer mención algunas de ellas como lo son:

- Máquinas lectoras para invidentes.
- Ayudas para los disminuidos físicos.
- Enseñanza asistida por medio de la computadora
- Terminales hablantes...

Es necesario mencionar que la implementación de las aplicaciones, se restringe a que exista un buen sintetizador, y este a su vez sea diseñado por medio de la concatenación de forma de onda, para que sea apto para dar una buena señal acústica casi como humana, es decir que se adaptable ó semejante a la pronunciación común del ser humano, como lo es la entonación, y expresión.

Por lo antes mencionado, se ha decidido crear una herramienta cuya finalidad es la búsqueda de palabras, frases u oraciones, esta herramienta es un sistema que genera audio a partir de un texto dado, a esto se le conoce como síntesis de voz.

Actualmente ya existen sistemas que realicen el proceso de traducir el texto a audio, pero en nuestro trabajo mejoraremos la calidad del proceso mencionado, como lo es mejorar el tiempo de búsqueda dentro del sistema.

1.2 Descripción general de los TTS

La conversión automática de un párrafo de texto a voz, pasa por varios procesos, como se mencionaba anteriormente, es indispensable tomar en cuenta las tareas de tipo lingüístico que son requeridas, como lo es el análisis del texto proporcionado, y así poder realizar una lectura natural. Para esto se genera una señal de voz la cual es proporcionada por procesos para la digitalización de la voz. Podemos representar como se lleva a cabo el proceso de sintetización de la siguiente manera:

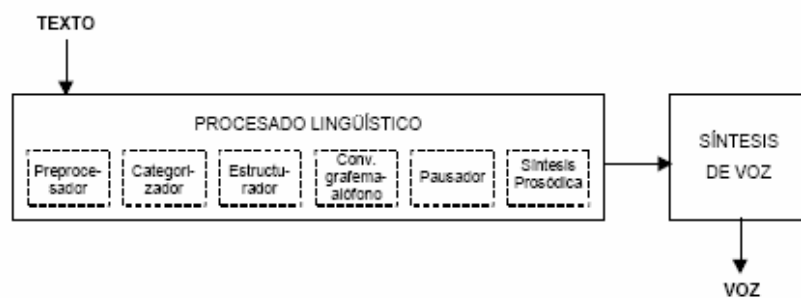


Fig. 1.1 Proceso de sintetización

1.3 Estado del problema.

Actualmente el sintetizador que existe, realiza esta función correctamente, pero se tarda bastante en buscar las palabras, frases u oraciones a sintetizar. Cuenta con un corpus de voz pregrabado el contiene la voz de hombre, en nuestro caso usaremos un nuevo corpus pregrabado, este ya incluye las dos voces, es decir, masculino y femenino.

Para solucionar el problema actual, en los capítulos siguientes definiremos principalmente lo que es síntesis de voz, así como los métodos que son usados con mayor frecuencia y los tipos de sintetizadores que actualmente existen y su desempeño o su funcionalidad a grandes rasgos. Se ira explicando paso a paso como se funcionan los sintetizadores y como es que se desarrolla nuestro sintetizador.

Existen dos tipos de ver la mejor manera para realizar un sintetizador, uno de ellos se conoce como “modelo de sistema”, este pretende interactuar a mayor frecuencia con el proceso de voz, esto ha dado paso a la existencia de dos tipos de sintetizadores que son articulatorios, y formantes.

El segundo tipo es el denominado “modelado de señal” y trata de interactuar mas con la señal de voz, que a diferencia del anterior este lo hace con el proceso de voz, y es aquí donde se crean los sintetizadores por concatenación, que en nuestro caso será el que usaremos, y que será explicado mas adelante.

1.4 Objetivo General

- Desarrollar un algoritmo capaz de reconocer la mejor opción en cuanto a la selección de unidad para síntesis de voz y probarlo.

1.4.1 Objetivos Específicos

- Investigación Bibliográfica sobre algoritmos existentes para síntesis de voz a través de la concatenación de unidades de longitud variable.
- Análisis de los métodos encontrados para seleccionar el más adecuado.
- Implementación del algoritmo.
- Pruebas de funcionalidad.
- Pruebas de inteligibilidad (calidad del sonido de las palabras y frases generadas).

1.5 Alcances y limitaciones

1.5.1 Alcances

- Creación de un algoritmo para síntesis de voz en español
- Implementación del algoritmo
- Creación de Estructura de Búsqueda
- Creación de una base de datos
- Implementación de una interfaz para la importación de los corpus.
- Implementación de la Estructura
- Mejoramiento de interface gráfica de usuario
- Mejoramiento del proyecto realizado con anterioridad

1.5.2 Limitaciones

- Se buscara la mejor implementación para la resolución del problema.
- La calidad del resultado de la implementación del sintetizador dependerá de la calidad de grabación del corpus proporcionado.
- El algoritmo creado e implementado deberá de ser capaz de resolver todas las peticiones que sean proporcionadas.
- Este trabajo asume que la información proporcionada ha sido comprobada, en cuanto a funcionalidad, así mismo que los corpus a usar están perfectamente grabados y etiquetados.

1.6 Software y Hardware

1.6.1 Software

- CSLU TOOLKIT creado por la Universidad de Colorado, nos ayudara para hacer el proceso de transmisión de señal de audio.
- MYSQL creación de una base de datos para manejar los corpus de voz.
- JAVA JSDK 1.4.2 compilador de java para la creación de las clases e interfaz
- Visual paradigm for UML Community Edition usado para la generación de diagramas UML.
- Visual Basic V6 usado para la implementación de una interfaz para la importación de los corpus.

1.6.2 Hardware

- Computadora personal con procesador Pentium 4 a 1.8 GHTZ
- Computadora personal portátil con procesador celeron a 1.8 GHTZ
- Computadora Sun con software para servidor Sun Ray, ambiente operativo Solaris 2.6