**UNIVERSIDAD NACIONAL DE COLOMBIA**

Facultad de Ingeniería

Departamento de Ingeniería de Sistemas y Computación

**ENCRIPTACIÓN DE TEXTO POR MEDIO DE ESPECTRO DE VOZ**

Integrantes:

Ferney Esteban Lizarazo López

Orlando Andrés Suarez Moreno

Bogotá

2015

**Objetivo general:**

Desarrollar un proyecto práctico relacionado con el procesamiento de mensajes de texto junto con el procesamiento de voz a través de un análisis espectral.

**Objetivos específicos:**

* Utilizar el proyecto para afianzar y aplicar los conocimientos teóricos adquiridos durante el curso de Teoría de la Información y Sistemas de Comunicación.
* Profundizar y analizar sobre los temas que no sean cubiertos en el curso y que sean necesarios para el proyecto.
* Aplicar las habilidades adquiridas en la herramienta MatLab a lo largo del curso para implementar el proyecto.

**Introducción:**

El presente proyecto, trata principalmente sobre el estudio para transformar señales de voz en llaves para acceso a textos previamente guardado y dirigido a un usuario o usuarios específicos que hayan procesado su voz en el sistema. Para el desarrollo del presente proyecto debemos abordar temas relacionados con práctica para la grabación de voz y adquisición de la señal producida, transformada de Fourier, procesamiento del espectro de voz, analizando este, por lo cual se pretende realizar una descomposición en partes simples que forman cada espectro, esto teniendo en cuenta su relación matemática con la transformada de Fourier, herramienta esencial para el análisis espectral.

La forma de interacción entre el sistema a desarrollar y el usuario es a través de la voz del mismo. La voz contiene mucha información de lo que se está diciendo y de la persona que lo esta diciendo.

**Justificación:**

Existen diversos campos de estudio que pueden ser abarcados en la ingeniería, estos con fundamentación en diversas teorías matemáticas, siendo la teoría de señales un área de importancia para la ingeniería de sistemas, dentro de la cual hay una cantidad enorme de campos de aplicación que permiten el trabajo incluso con diferentes áreas ya manejadas en el transcurso de la carrera, por medio del presente proyecto se tiene como principal motivación la aplicación de los temas tratados a un ejercicio práctico en el cual se realice un mejor conocimiento y compresión a través del desarrollo y presentación de este.

Estado del arte:

En la actualidad, las contraseñas proporcionan algo de protección, pero recordar y saber dónde están guardados los diferentes códigos de cada máquina es un problema en sí mismo.

Matemáticamente el análisis espectral está relacionado con una herramienta llamada transformada de Fourier. Dada una señal o fenómeno ondulatorio de amplitud \scriptstyle s(t) esta se pude escribir matemáticamente como la siguiente combinación lineal generalizada:

s(t) = \int_\R A(\nu)e^{-2\pi i\nu t}d\omega 

Es decir, la señal puede ser concebida como la transformada de Fourier de la amplitud \scriptstyle A=A(\nu). Ese análisis puede llevarse a cabo para pequeños intervalos de tiempo, o menos frecuentemente para intervalos largos.

El acceso físico directo a habitaciones y dispositivos también se puede asegurar acoplando reconocimiento de voz con sistemas de apertura de puertas. Los terminales de salidas en los aeropuertos serán capaces de procesar a los pasajeros de manera más rápida. Los automóviles, la maquinaria de construcción, los barcos y los aviones también estarán protegidos contra robos.

**Teoría de funcionamiento**

**Fundamentos de la voz:**

La voz es una onda de presión acústica que es generada voluntariamente a partir de movimiento de la estructura anatómica del sistema fonético humano. La producción de esta arranca en el cerebro con la conceptualización de la idea a transmitir, asociada a una estructura lingüística, realizando una selección de palabras y orden de acuerdo a unas reglas gramaticales. Posteriormente el cerebro produce los comandos nerviosos que se encargan del movimiento de los órganos vocales para la producción de sonido.

La unidad mínima de una cadena hablada es el fonema, este posee un carácter distintivo en la estructura de la lengua. Varios fonemas dan silabas, y estas a su vez conforman palabras, las cuales llegan a formar oraciones.

Los parámetros principales del sistema articulatorio son: cuerdas vocales, paladar, lengua, dientes, labios, y mandíbulas. Desde la ingeniería es lógico pensar que el sistema físico puede verse como un filtro, cuya función de transferencia depende del sonido articulado y, por ende, de la posición de los órganos que intervienen en la producción de la voz. La entrada del filtro puede modelarse a través de una señal de excitación, que consta del paso del aire generado por los pulmones a través de la tráquea y las cuerdas vocales, lo cuales es dependiente del sonido generado.

La frecuencia de la voz depende de varios factores, como el tamaño y masa de las cuerdas vocales, la tensión de estas cuerdas, la tensión que se le aplique a estas y de la velocidad del flujo de aire que viene de los pulmones. A mayor tamaño, menor frecuencia de vibración de estas cuerdas, y a mayor tensión aplicada, la frecuencia aumenta, dando los sonidos más agudos. La frecuencia también presenta aumento al aumentar la velocidad del flujo de aire.

**Reconocimiento automático del habla:**

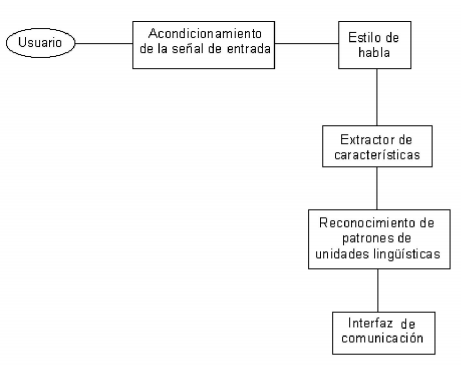
**Tipos:**

* Aplicaciones locales: en esta área se busca presentar interfaces hombre-máquina, que sustituyan la utilización de teclado y mouse, dotando al usuario de una movilidad que estos dispositivos le restan, así mismo facilidad el uso de máquinas a usuarios discapacitados.
* Respuesta vocal interactiva: se involucra la difusión o captura de información por un gran número de usuarios, tales como mensajería vocal, consulta de cuentas bancarias, trasmisión de información general, entre otros.
* Automatización de sistemas telefónicos: marcación por voz, manejo de agendas, directorio público, entre otras.

**Principales áreas de trabajo en reconocimiento de voz**

* Procesamiento de la señal de voz
* Técnicas de reconocimiento de patrones
* Diferentes estilos de habla
* Dependencia del locutor
* Tarea de reconocimiento
* Bases de datos para entrenamiento y reconocimiento

Diagrama básico de reconocimiento del habla, en el diagrama se muestran algunos puntos importantes para la realización de la implementación del sistema.



**Procesamiento de la señal de voz**

La primera tarea en un reconocimiento de voz es el procesamiento de la señal de voz de entrada al sistema, con el fin de extraer la información acústica relevante para la autenticación a realizar. Los rasgos o características a extraer de la señal de voz, son el resultado de un largo proceso de investigación sobre procedimientos de parametrización de la voz.

Llegando a una solución actual mayormente extendida, una parametrización de la envolvente espectral que incluye consideraciones preceptúales (sensación interior) a partir del funcionamiento del oído.

Se selecciona un subconjunto con los parámetros más distintivos.

La señal que se transmite como señal de entrada puede venir acompañada de efectos perturbadores, los cuales deben ser eliminados en lo posible, para lo cual se tienen en cuenta tres aspectos principales:

* Detección robusta de voz: procedimientos de selección entre voz o ruido (silencio) para distintos tipos de ruido.
* Reducción de ruido: procedimientos que actúan en la señal de voz, buscando compensar el efecto del ruido sobre la parametrización de la voz.

**Obtención de información mediante micrófono**

El micrófono es un transductor electro acústico. El audio es un fenómeno analógico. Para la grabación de una señal de voz se realiza una conversión de la señal analógica del micrófono en una señal digital por medio del conversor A/D en la tarjeta de sonido.

Hay dos factores importantes durante este proceso. La taza de muestreo o que tan seguido los valores de voltaje son grabados. Segundo, son los bits por segundo, o que tan exactamente los valores son grabados. Un tercero podría ser el número de canales (mono o estereo), para las aplicaciones de reconocimiento de voz un canal mono es suficiente.

La mayoría de aplicaciones vienen con valores pre-determinados. Dado a que el habla es relativamente de bajas frecuencias (entre 100Hz-8kHz), una frecuencia de muestreo de 16000 muestras/seg provee una mayor exactitud en la adquisición de la información, la frecuencia de nyquist. La cuantificación más comúnmente usada, es de 8 bits, mínimo requerido para una calidad baja, se obtienen excelentes resultados aumentando la cuantificación a 16 bits.

A partir de la señal analógica obtenida se hace necesario convertir la señal a formato digital para poder procesarla en la computadora lo que se realiza mediante dos procesos: muestreo y cuantificación. La señal vocal tiene componentes de frecuencia que pueden llegar a los 10 kHz., sin embargo la mayor parte de los sonidos vocales tiene energía espectral significativa hasta los 5 Khz. Para el sistema se puede optar por una frecuenta de muestreo de 44100Hz y una resolución de 16 bits mono canal.

**Pre procesamiento:**

Para el análisis es necesario realizar un pre-procesamiento de la señal vocal. Esto se realiza a través de técnicas que permitan extraer la información acústica directamente a partir de la señal vocal emitida.

* Filtro: se realiza un filtrado de la señal para reducción de efectos del ruido que acompaña la señal de voz que es ingresada. Se deberá considerar que la elección de frecuencias de corte y paso, jugaran un papel importante en la distinción de voces espectralmente parecidas.
* Segmentación (puntos de inicio y fin de la señal para la eliminación del ruido) se debe determinar adecuadamente estos puntos, diferenciando partes de la señal que llevan voz de aquellas que no, este procedimiento evita el gasto de memoria y tiempo de cálculo.

**Parametrización:**

Cepstrum: los sonidos de la voz se pueden representar mediante un espectrograma, que indica los componentes de la frecuencia de la señal de voz. El espectro nos proporciona información acerca de los parámetros del modelo de producción de voz, tanto de la excitación como del filtro que representa el tracto vocal.

**Modelo de obtención de los coeficientes cesptrales**

En el sistema de reconocimiento de voz en MATLAB existe una función para obtener los coeficientes cepstrales utilizando la FFT. La función utilizada es la rceps, que nos proporciona el cepstrum real de la función ingresada, por medio del algoritmo mostrado en la figura de arriba. La razón principal para utilizar los coeficientes cepstrales es que tienen la ventaja adicional que uno puede derivar de ellos una serie de parámetros que son invariantes sin importar las distorsiones que puedan ser introducidas por el micrófono o por cualquier sistema de transmisión

**Propuesta:**

Implementar en Matlab un sistema que registre usuarios con algunos datos, su de voz y en lo posible un mensaje con destinatarios en específico, los usuarios tendrán acceso a estos mensajes a través de una autenticación por voz, esto por fines de privacidad, ejemplificando situaciones donde se podría aplicar, siendo la biometría por voz, una solución a problemas que enfrentan tradicionales sistemas de autenticación. Se daría al usuario un proceso de autenticación sin mayor esfuerzo, con el fin de no tener fraudes.

Según datos consultados, 49% de los usuarios afirman que una autenticación lleva demasiado tiempo, 67% de usuarios de móviles restablecen una contraseña por lo menos una vez al mes, y el 85% de usuarios sienten frustración con sistemas actuales de autenticación.

Se hará la utilización de una voz de huella única del usuario para la autenticación, la cual será de modo pasivo (el usuario dice cualquier cosa y se verifica con la base de datos) o según los resultados y el tiempo se podría llegar a un modo activo en donde la persona deberá decir una frase, de igual forma dependerá del modo pasivo.

Entonces lo que se hará es un procesamiento digital de las señales de voz. El procesamiento digital de señales de voz tiene una gran variedad de aplicaciones, existe una base para el tratamiento digital de señales, que puede ser implementada para lograr obtener lo que nos interese según la aplicación. El Sistema de Reconocimiento de Voz es una de las aplicaciones del procesamiento digital de señales de voz. El sistema consiste en obtener una señal de voz que permita reconocer qué palabra se está hablando y quien la está realizando (dentro de un número limitado de locutores).

La biometría de voz analiza numerosas características únicas en la voz de cada persona, realizando comparaciones con registros almacenados en una base de datos.

Se dará una interfaz acorde con parámetros de diseño y usabilidad amigables para el usuario, esto también implementado en la herramienta Matlab que da una gran cantidad de herramientas para abordar los fines propuestos.

**Bibliografía.**

* <http://www.nuance.es/landing-pages/products/voicebiometrics/>
* Algoritmos y métodos para el reconocimiento de voz en español mediante sílabas. <http://www.saber.cic.ipn.mx/cake/SABERsvn/trunk/Repositorios/webVerArchivo/277/3>
* <https://prezi.com/egzjauff07gu/analisis-espectral-usando-matlab/>
* <http://es.wikipedia.org/wiki/Se%C3%B1al_de_voz>
* <http://es.slideshare.net/ABEL170/analisis-espectral-en-matlab-7757614>
* <http://es.wikipedia.org/wiki/Espectro_de_frecuencias>
* <http://www.nuance.es/empresas/solucion/soluciones-de-atencion-al-cliente/servicios-y-soluciones/soluciones-de-recepcion-de-llamadas/software-biometrico-de-autenticacion-de-voz/index.htm>
* <https://www.dspace.espol.edu.ec/bitstream/123456789/10740/11/MATLAB_GUIDE.pdf>
* <http://www.coimbraweb.com/documentos/comunicacion/p1_analisisespectral.pdf>
* <http://www.frsn.utn.edu.ar/tecnicas3/problemas/Reconocimiento%20de%20voz.pdf>