

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR SEDE AMBATO

SERÉIS MIS TESTIGOS

ESCUELA DE INGENIERIA DE SISTEMAS

TEMA:

"HUMANOIDE INTELIGENTE QUE INTERACTÚE CON MULTIUSUARIOS MEDIANTE COMANDOS DE VOZ"

DISERTACION DE GRADO PREVIO A LA OBTENCION DEL TITULO DE INGENIERIA DE SISTEMAS Y COMPUTACIÓN

AUTOR:

PABLO ISRAEL MORALES PAREDES

DIRECTOR:

ING. MSC. JANIO JADAN

BIBLIOTECA

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR SEDE AMBATO

BIBLIOTECA

Ambato – Ecuador
Febrero 2009

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR SEDE AMBATO

HOJA DE APROBACIÓN

Tema:

"Humanoide inteligente que interactúe con multiusuarios mediante comandos de voz"

Autor:

PABLO ISRAEL MORALES PAREDES

Janio Jadán, Ing. MSc.
DIRECTOR DE LA DISERTACIÓN.

f. _____

Pablo Montalvo. Ing. MSc.
CALIFICADOR

f. _____

Marco Polo Silva, Ing. MSc
CALIFICADOR

f. _____

Santiago Acurio, Ing.
DIRECTOR UNIDAD ACADÉMICA

f. _____

Pablo Poveda Mora, Ab.
SECRETARIO GENERAL PUCESA

f. _____

SECRETARIA GENERAL
PROCURADURIA

DECLARACIÓN DE AUTENTICIDAD Y RESPONSABILIDAD

Yo, Pablo Israel Morales Paredes portador de la cédula de ciudadanía N°. 180353682-8 declaro que los resultados obtenidos en la investigación que presento como informe final, previo la obtención del título de Ingeniero de Sistemas son absolutamente originales, auténticos y personales.

En tal virtud, declaro que el contenido, las conclusiones y los efectos legales y académicos que se desprenden del trabajo propuesto de investigación y luego de la redacción de este documento son y serán de mi sola y exclusiva responsabilidad legal y académica.

Pablo Israel Morales Paredes

CI. 180353682-8

Agradecimiento

Principalmente a Dios por la vida que me ha dado y la sabiduría necesaria para afrontar cada obstáculo y tomar decisiones importantes en mi vida para cumplir todos mis objetivos y mis sueños.

A mi Madre, la cual ha sido mi fortaleza y mi guía principal ya que ella ha sido el pilar fundamental que con su sacrificio, esfuerzo y ejemplo me ha ayudado a para edificarme como persona.

A mis maestros ya que ellos me transmitieron todos sus conocimientos profesionales, experiencias y sobretodo su gran amistad.

Y también quiero dar gracias a todas las personas que se involucraron directa e indirectamente en el transcurso del desarrollo del proyecto, mil gracias.

Dedicatoria

A mis Padres

Dedico con mucho cariño y afecto este Proyecto de Tesis ya que fueron quienes me han sabido guiar por el camino del bien al mismo tiempo me inculcan para que cada día me prepare siendo mi gran apoyo moral, a la vez que me ayudaron a cumplir uno de mis grandes anhelos.

A mi Abuelita

Quien ha sido un mi gran ejemplo de lucha y superación, ya que con ese gran deseo de vivir me ha enseñado que tengo que ser fuerte ante adversidades de la vida.

A mis Hermanos

Con quienes comparto todos los momentos gratos de mi vida incentivándome a que día a día me supere.

Resumen

Este proyecto de tesis consiste en brindar información a través de un robot inteligente que reconoce la voz de las personas y va a emitir una respuesta dependiendo del contenido o tema de la pregunta. De esta manera el robot va a entablar una conversación puntual de temas relacionados con la Carrera de Ingeniería de Sistemas y como contiene una base de datos que en el futuro se estima va a contener información de toda la Universidad. El robot en el cual se desarrolla este proyecto de tesis inicialmente funcionaba con botones pre programados con información acerca de la Universidad, pero en esta ocasión el paso gigante que se dio es que el robot ya va a reconocer la voz de las personas mediante un software especializado en ese campo. Además del software mencionado, se trabaja con una base de datos que va a almacenar la información con la que se va a manejar el robot. Para poder hablar con el robot se necesita un micrófono el cual va ser el dispositivo necesario y fundamental en este proyecto. Para la utilización de este robot primero se necesita que esté conectado a la corriente eléctrica mediante el cable de poder, luego presionar el primer botón para el encendido que se encuentra en la parte derecha del robot, una vez hecho este proceso va a inicializar la aplicación del robot la cual dará inicio a la interacción del usuario con la maquina mediante una conversación. Es así que este proyecto va a facilitar información a los usuarios además de estar relacionados directamente con la tecnología actual que esta creciendo a pasos agigantados con el propósito de eliminar tareas que para los seres humanos resultarían repetitivas y exhaustas.

Abstract

This thesis Project allows an intelligent robot to give information by recognizing people´s voice and gives back an answer depending on the information or the topic of the question given to it. That way the robot begins a fluent conversation based on clear topics about the career of system engineering and in the future we hope to include more information on the data base of the robot in order to allow it to have these same results with other careers in the university.

At the beginning the robot was programmed to give pre recorded information by just a touch of a button but a bigger step was made when now the robot is allowed to recognize people's voice using software made for this field. In addition to this software it works with database installed that improves as the robot saves additional information. To speak with the robot a microphone is required and necessary in this project. The robot must be connected to the electrical outlet using a power cord then the first button is pressed on at the right side of it. Once this is done the interrelation among the robot and the person begins in a conversation. So this project has become a great achievement not only because it allows users to be acquainted with technology but also will stop repetitive and boring tasks given to humans.

TABLA DE CONTENIDOS

DECLARACIÓN DE AUTENTICIDAD Y RESPONSABILIDAD iii
Agradecimiento iv
Dedicatoria v
Resumen vi
Abstract vii
TABLA DE CONTENIDOS viii
TABLA DE GRÁFICOS x
FIGURAS x
TABLAS xi
Introducción 1
CAPITULO I. 3
PROYECTO DE LA INVESTIGACIÓN 3
1.1. Antecedentes 3
1.2. Significado del Problema 4
1.3. Definición del Problema 4
1.4. Planteamiento del tema 5
1.5. Delimitación del tema 5
1.6. Hipótesis 7
1.7. Objetivos 8
1.7.1. General 8
1.7.2. Específicos 8
1.8. Metodología de Trabajo 9
1.8.1. Descriptiva: 10
1.8.2. Exploratoria: 11
1.9. Justificación 11
CAPITULO II. 13
MARCO TEORICO 13
2.1. Robótica 13
2.1.1 Automatización 14
2.1.2. Tareas Programadas 16
2.1.3. Operaciones de Proceso 16
2.2. Bases de Datos 17
2.2.1. Almacenamiento 18

2.2.2. Gestión de la BD 20
2.3. Electrónica 21
2.3.1. Tarjeta de Reconocimiento de Voz 22
2.3.1.1. Características 23
2.3.1.2 Funcionamiento 23
2.3.1.3. Configuración 23
2.4. Inteligencia Artificial 25
2.4.2. Reglas 26
2.4.2.1. Motor de inferencia o intérprete de reglas 26
2.4.2.2. Base de conocimiento 27
2.4.2.3. Base de afirmaciones o memoria de trabajo 27
2.4.2.4. Interfaz de usuario 27
2.4.2.5. Base de datos 27
2.4.3. Algoritmos de Búsqueda 28
2.4.4. Reconocimiento de Voz 28
2.4.4.1. Diseño de un Sistema de Reconocimiento de Voz 29
2.4.4.2. Aprendizaje 29
2.4.4.3. Modelo del lenguaje 31
2.4.4.4. Hardware – Características 32
2.4.4.5. Software – Características 38
2.4.4.6. Usos y Aplicaciones 39
2.4.5. Procesamientos de Lenguajes Naturales 42
2.4.5.1. Dificultades en los PLN 43
2.4.5.1.1. Ambigüedad 43
2.4.5.1.2. Detección de separación entre las palabras 44
2.4.5.1.3. Recepción imperfecta de datos 45
2.4.5.1.4. Componentes 45
CAPITULO III. 47
DESARROLLO DEL PROYECTO 47
3.1. Enfoque 47
3.2. Metodología de Desarrollo de Software 49
3.2.1. Obtención de requisitos 49
3.2.2. Diseño Global del prototipo 50
3.2.3. Construcción del prototipo 53

BIBLIOTEC

3.2.4. Pruebas para verificar que cumpla con los requisitos para los que fue diseñado y mejoramiento del prototipo 58
3.3. Selección de Información 59
3.4. Procesamiento de la Información 60
3.5. Software utilizado en la Implementación 61
3.5.1. Dragon Naturally Speaking 62
3.5.2. Microsoft Visual Basic 6.0 63
3.5.3. Microsoft Access 64
3.6. Tipos de Usuarios 64
3.7. Análisis de la Estructura de los Datos 65
3.8. Comandos de Voz 67
3.9. Analizador Léxico 68
3.10. Herramientas de Interacción y Voz 72
3.10.1. VSA (Visual Servo Automation) 73
3.10.2. Dragon Naturally Speaking 75
3.10.2.1. Creación de un Usuario 76
3.10.2.2. Centro de Precisión 79
3.11. Aparato Fonador y medios de Interacción 82
3.12. Dialogo 83
3.13. Control 83
CAPITULO IV. VERIFICACION Y VALIDACION DE RESULTADOS 87
4.1. Análisis General 87
4.2. Pruebas del Proyecto de Tesis 88
4.2.1. Prueba de Caja Negra 88
4.2.2. Grafos de Causa – Efecto 90
CONCLUSIONES 92
RECOMENDACIONES 93
GLOSARIO DE TERMINOS 94
BIBLIOGRAFIA 100
ANEXO MANUAL DE USUARIO

TABLA DE GRÁFICOS

FIGURAS

Figura 1: Proceso de reconocimiento de voz 10

Figura 2 Diagrama de los componente de una tarjeta de sonido 32
Figura 3: Esquema de colores en la tarjeta de Sonido Sound blaster 37
Figura 4 . Interacción Usuario - Máquina 51
Figura 5. Ventana Principal del Programa 52
Figura 6. Caja de Herramientas 54
Figura 7. Ventana de Componentes 55
Figura 8. Microsoft Agente 55
Figura 9. Consola VSA en el Formulario del Programa 56
Figura 10. Formulario Principal del Programa 57
Figura 11. Secuencia de Interacción Maquina – Usuario 60
Figura 12: Merlín Figura 13: Genio 72
Figura 14: Entorno del VSA 74
Figura 15. Asistente de Dragon 77
Figura 16. Barra Principal de Dragon 78
Figura 17. Ventana de Inicio del Entrenamiento de Usuario 80
Figura 18. Ventana para añadir palabras nuevas 81
Figura 19. Ventana de entrenamiento de palabras 81
Figura 20. Grafo Causa - Efecto 90

TABLAS

Tabla 1. Resumen de la interacción humano – máquina 58
Tabla 2: Estructura de la Base de Datos COMUNES 66
Tabla 3: Estructura de la Base de Datos CONOCIMIENTO 66
Tabla 4: Código Fuente de la extracción de tokens (palabras relevantes) ... 71
Tabla 5. Tabla de Causas 90
Tabla 6. Tabla de Efectos 91

Introducción

En la actualidad existen una amplia gama de sistemas que brindan al usuario información de una determinada área, sistemas que generan como resultado datos generales, los cuales son de gran importancia para la persona que requiera esta información. Es así que cada día los sistemas tratan de guardar información primordial y básica proporcionando resultados claros que ayuden al conocimiento y desarrollo intelectual de las personas.

La mejor manera de adquirir conocimiento fundamental y rápido, es la comunicación entre personas, intercambiando información específica y puntual; hoy en día esta comunicación es cada vez más amplia debido al uso de las nuevas tecnologías de la comunicación e Información (TICS). Esto da a lugar a que las personas nos comuniquemos por medio de máquinas e incluso con ellas. Con el avance tecnológico hoy en día se desarrollan nuevos dispositivos que automaticen el proceso de comunicación, y un área en particular es el desarrollo de robots, los cuales emplean mecanismos que ayudan a la intercomunicación entre la persona (usuario) y el mecanismo inteligente, que tendrá información básica, fundamental o completa dependiendo el ámbito a satisfacer.

El presente proyecto se enfoca al área de la Robótica, en donde se pretende utilizar el autómata programable existente en la Escuela de Ingeniería de Sistemas para ampliar su funcionalidad de comunicación. Mediante técnicas de reconocimiento de voz y hardware específico se

permitirá el reconocimiento de voz de los usuarios, dando lugar a una interacción hombre-máquina con el soporte de técnicas de Inteligencia Artificial.

Al finalizar el proyecto, la PUCESA contará con un autómata "inteligente" que pueda brindar información institucional a cualquier persona independientemente de las horas de oficina.

CAPITULO I.

PROYECTO DE LA INVESTIGACIÓN

1.1. Antecedentes

Para este proyecto se tiene como referencia una investigación desarrollada anteriormente la cuál permite a un robot ejecutar una secuencia de órdenes cuando la persona se acerque a su ambiente, el mismo que es controlado por medio de sensores que ayudan a saber la presencia de alguien que quiera interactuar con él. Por lo tanto esta investigación nos da una pauta para continuar en este difícil pero no imposible mundo de los robots tratando de implementar mecanismos para poder establecer una comunicación hombre - máquina. Es interesante ampliar este tema ya que de alguna u otra forma se impulsa el desarrollo de sistemas y mecanismos inteligentes que ayuden al ser humano en los trabajos más difíciles ya sea realizando tareas repetitivas que como fuentes de información externas provean órdenes o instrucciones que sean entendidas por estos dispositivos.

1.2. Significado del Problema

El proyecto significa un trabajo muy importante si queremos dar información atractiva, clara y detallada de la Escuela de Ingeniería de Sistemas. Nosotros podemos dar información rápida y general al que lo necesite por lo tanto recurren a los trípticos, es así que nosotros nos enfrentamos a los siguientes problemas:

- Información general de la carrera.
- Nadie nos puede dar información constante de la carrera.
- No existe un incentivo tecnológico que llame la atención además de interactuar y simular ser un ser humano de tal manera que se incentive nuevos estudiantes ser parte de esta interesante carrera.
- No hay quien de información de la carrera cuando no se encuentra la secretaria de la Escuela.

1.3. Definición del Problema

Debido a la carencia de información detallada y de un Humanoide Inteligente que interactúe con multiusuarios en la Escuela de Sistemas, se ha visto oportuno el desarrollo de este proyecto de tesis, el cuál va a brindar información fundamental de la carrera conectándose con el usuario mediante comandos de voz, creando así interés en las personas y a la vez estos sean protagonistas directos vinculándose con el robot a la tecnología actual.

1.4. Planteamiento del tema

El tema de este proyecto de tesis es:

"Humanoide Inteligente que interactúe con multiusuarios mediante comandos de voz"

1.5. Delimitación del tema

Como en todo proyecto, es necesario definir el tiempo de desarrollo. En este caso el tiempo estimado para este proyecto es de 9 meses entre desarrollo y mejoras, teniendo en cuenta que la programación de un ente inteligente tiene un cierto grado de dificultad.

El prototipo funcionará en las instalaciones de la Escuela de Sistemas de la PUCESA, dando facilidad a las personas que deseen adquirir información de la misma.

Preguntas Básicas

¿Qué va hacer el humanoide inteligente?

Se tiene claro que el robot va a responder a cierta orden que nosotros le demos por medio hablado, el mismo que nos va a dar una respuesta con información específica de la carrera, sin embargo la forma en que el robot lo reciba puede resultar un proceso muy complejo.

¿Cómo lo va hacer?

Se va a llenar de información acerca de la carrera en una base de datos, la cual nos va a ayudar a comparar con el comando vocalizado o sentencia ingresada por el usuario la misma que es analógica, por lo tanto hay que convertir las palabras o sentencias a digitales para que las computadora nos entienda, es aquí donde toman parte fundamental el software de comandos de voz, el mismo que va reconocer un vocabulario definido de palabras. Es así como el sistema nos va a proveer de información específica en respuesta a la sentencia ingresada anteriormente por el usuario.

¿Para qué implementar el humanoide?

La respuesta simplemente es ahorrar tiempo a la secretaria de la carrera además de relacionarnos directamente con la tecnología y comprender el desarrollo del mismo. Ayudando así las personas interesadas por la carrera de la universidad y alguna forma ayudar a los usuarios con discapacidades físicas limitados en su habilidad para utilizar un teclado, a comunicarse de mejor manera con este tipo de robot. Por esta sola razón, es esencial el

desarrollo continuo del reconocimiento de voz en post de un mejor entendimiento hombre máquina.

1.6. Hipótesis

El humanoide inteligente logrará reconocer la voz de varios usuarios y dará respuesta a una pregunta o sentencia específica.

VARIABLES

Variable Dependiente

Cantidad de multiusuarios que interactúen con el humanoide

Variable Independiente

Mecanismo de información interactivo de los usuarios de Ingeniería de Sistemas de la PUCESA.

1.7. Objetivos

1.7.1. General

Implementar un sistema de Reconocimiento de Voz que permita al humanoide de la Escuela de Ing. de Sistemas de la PUCESA reconocer la voz humana.

1.7.2. Específicos

- Dar información de la carrera por medio del humanoide inteligente de forma permanente y oportuna.

- Permitir a las personas se relacionen directamente con la tecnología.

- Incentivar a los estudiantes al desarrollo de dispositivos inteligentes que serán de gran ayuda en el futuro.

1.8. Metodología de Trabajo

El reconocimiento de voz permite que las computadoras tengan la capacidad para comprender el lenguaje hablado de los usuarios y una vez entendido puedan realizar funciones específicas ordenadas por el usuario.
Por lo tanto el proceso que involucra el interactuar con el humanoide es:

Comunicación: Envía lo reconocido a la aplicación y ésta se encarga de la comunicación entre usuarios y máquina. El proceso se muestra en la **Figura 1.**

- **Procesamiento:** Convierte las entradas de voz a una forma que el microcontrolador pueda procesar, es decir, convertir la señal analógica a digital.

- **Reconocimiento:** Identifica lo que se dijo o se ingresó de forma hablada.

- **Programa de Aplicación:** Controla el software de reconocimiento de voz además de controlar la interacción de movimientos e información con el usuario

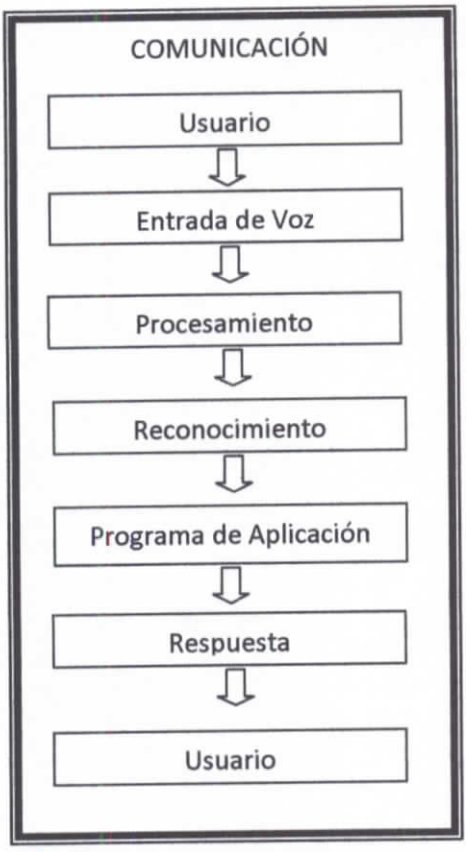


Figura 1: Proceso de reconocimiento de voz

De esta manera los tipos de metodología a usar en esta investigación son:

1.8.1. Descriptiva:

Ya que es necesario el detalle del tema, tomando en cuenta la información que involucra y los medios necesarios de desarrollo.

1.8.2. Exploratoria:

Se necesita saber información preliminar que ayude en la investigación y desarrollo aclarando los conceptos de este tema ya que su contenido es amplio y se necesita parámetros validos para su avance y la interrelación con los usuarios.

1.9. Justificación

Este Proyecto de Tesis va a desempeñar un papel muy importante en la accesibilidad automática e interactiva con los usuarios de la PUCESA, especialmente con los alumnos de Ingeniería de Sistemas, los cuales accederían a la información de la carrera en forma directa y oportuna, muchas de las veces necesitamos más información que los trípticos no lo tienen y su contenido es muy reducido, es aquí donde sería un factor muy importante el humanoide ayudando a proveer información precisa en respuesta a las necesidades del usuario.

Además de ser un dispositivo que contiene información útil de la carrera, físicamente se acerca al entorno humano, siendo amigable a las personas quienes visitan las instalaciones de la Universidad o a su vez también exponer externamente el desarrollo y utilidad del mismo ya sea en exposiciones en otras Universidades Locales o Nacionales creando

expectativa en el desarrollo de proyectos y robots inteligentes que interactúen con los seres humanos.

Es así que al tratar de simular algunas características de los seres humanos se ha creado un mecanismo único que va a tratar de resolver un problema específico en un área específica en un tiempo específico. En este caso el robot va a facilitar información de forma amigable y confiable al usuario.

CAPITULO II.

MARCO TEORICO

2.1. Robótica

La robótica es una rama muy importante dentro de la tecnología, ya que nos permite crear un prototipo capaz de ayudar a las personas en situaciones peligrosas y también las tediosas tareas repetitivas. Las ciencias y tecnologías de las que deriva la robótica podrían ser: el álgebra, los autómatas programables, las máquinas de estados, la mecánica, la electrónica y la informática. Siendo primordial en la actualidad como una ayuda significativa en los diferentes campos industriales y científicos.

Clasificación de los robots

A continuación se detallan los diferentes tipos de robots que se han desarrollado hasta la actualidad, los cuales se clasifican de la siguiente manera:

1. **Robots Play-back,** los cuales regeneran una secuencia de instrucciones grabadas, como un robot utilizado en recubrimiento por spray o soldadura por arco. Estos robots comúnmente tienen un control de lazo abierto.

2. **Robots controlados por sensores,** estos tienen un control en lazo cerrado de movimientos manipulados, y hacen decisiones basados en datos obtenidos por sensores.

3. **Robots controlados por visión,** donde los robots pueden manipular un objeto al utilizar información desde un sistema de visión.

4. **Robots controlados adaptablemente,** donde los robots pueden automáticamente reprogramar sus acciones sobre la base de los datos obtenidos por los sensores.

5. **Robots con inteligencia artificial,** donde los robots utilizan las técnicas de inteligencia artificial para hacer sus propias decisiones y resolver problemas.

2.1.1. Automatización

¿Que es un sistema automatizado?

Es un dispositivo o conjunto de reglas que realizan un encadenamiento automático y continuo de operaciones capaces de procesar una información de entrada para producir otra de salida.

Se puede decir también que "La automatización es un sistema donde se trasfieren tareas de producción, realizadas habitualmente por operadores humanos a un conjunto de elementos tecnológicos". [1]

Un sistema automatizado consta de dos partes principales:

- Parte Operativa
- Parte de Mando

La *Parte Operativa* es la parte que actúa directamente sobre la máquina. Son los elementos que hacen que la máquina se mueva y realice la operación deseada. Los elementos que forman la parte operativa son los accionadores de las máquinas como motores, cilindros, compresores .y los captadores como fotodiodos, finales de carrera.

La *Parte de Mando* suele ser un autómata programable (tecnología programada), aunque hasta hace bien poco se utilizaban relés electromagnéticos, tarjetas electrónicas o módulos lógicos neumáticos (tecnología cableada). En un sistema de fabricación automatizado el autómata programable está en el centro del sistema. Este debe ser capaz de comunicarse con todos los constituyentes de sistema automatizado.

---

1 http://www.sc.ehu.es/sbweb/webcentro/automatica/WebCQMH1/PAGINA%20PRINCIPAL/Automatizacion/Automatizacion.htm

2.1.2. Tareas Programadas

Se entiende por mecanización la incorporación de máquinas para ejecutar determinadas tareas; con el propósito de realizar, de forma automática, una serie de acciones que anteriormente se llevaban a cabo de modo manual. El resultado final es una mayor rapidez en el trabajo y, en determinados casos, una mejor calidad en el resultado.

La automatización consiste en eliminar, de forma parcial o total, la intervención de personas en la ejecución del trabajo. Las máquinas automáticas pueden repetir continuamente la acción para la que están diseñadas, pero sin posibilidad de variar su funcionamiento. Las consecuencias más directas son el aumento de la productividad y la realización de tareas desagradables o peligrosas sin participación humana.

2.1.3. Operaciones de Proceso

Estas operaciones permiten que los robots tengan similitud con las actividades cotidianas del ser humano, ya sea programándolos con procesos fáciles para las personas y agregándoles características especiales para que sea una ayuda fundamental en las tareas difíciles y a la vez que tenga autonomía desarrollando de esta manera la Inteligencia Artificial. Según Ollero "Los robots autónomos o inteligentes son los más evolucionados desde el punto de vista del procesamiento de información. Son maquinas

2.2.2. Gestión de la Base de Datos

La Gestión de una base de datos es un sistema desarrollado que hace posible acceder a datos integrados que atraviesan los límites operacionales, funcionales u organizacionales de una empresa.

Objetivos en el uso de un sistema de gestión de base de datos:

- Oportunidad, asociado a la eficiencia y eficacia.
- Disponibilidad, permitiendo la accesibilidad de datos
- Consistencias (oportunidad + disponibilidad), como calidad de datos
- Evolución, para adaptarse al entorno
- Integridad, en el nivel de los datos así como el sistema.

Objetivos del sistema de gestión de base de datos que podemos identificar son:

- Independencia de datos
- Accesibilidad limitada
- Datos al día y sin redundancias
- Consistencia
- Interfaz única
- Entrada directa a los datos
- Recuperación por diferentes accesos

- Función completa de interrogantes
- Estandarización
- Seguridad

2.3. Electrónica

La electrónica es la parte esencial de la robótica la cual permite manipular técnicamente a los elementos que la conforman por lo tanto se la ha denominado como arte ya que utiliza dispositivos en los que se mueven cargas eléctricas, para obtener, manejar y utilizar información. La electrónica es un instrumento físico además de ser una herramienta de cálculo en las aplicaciones más modernas como son los ordenadores digitales y automatismo.

"Debido a la variedad de técnicas que se aplican junto con la electrónica o junto con las cuales se aplica la electrónica, la utilización última del total no es exclusivamente electrónica. La acertada utilización de técnicas electrónicas y no electrónicas para obtener un resultado útil constituye la Ingeniería en Sistemas. En principio, puede haber sistemas que no tengan ninguna parte electrónica, si bien en la vida moderna casi todos los sistemas complejos tienen parte electrónica". [González Bernaldo de Quiros, 1983][4]

[4] J. Gonzalez B. "Electronica General Tomo V". España; 1983

Por tal motivo la electrónica es la parte fundamental del desarrollo de la robótica y especialmente de este proyecto de tesis ya que se necesitan componentes electrónicos programables para las tareas especificas como movimientos y como se citó anteriormente se necesita de cargas eléctricas para manejar y utilizar información que va a contener el ente inteligente.

2.3.1. Tarjeta de Reconocimiento de Voz

Algunas tarjetas adaptadoras de audio están equipadas con software que es capaz de reconocer tonos de voz.

También puede lograr el reconocimiento de voz a partir de su adaptadora actual en la forma de un complemento de software. El reconocimiento de voz, como su nombre implica, es cuando a su computadora se le "enseña" a reconocer formas de palabras habladas y reaccionar a ellas.

"Por lo general, los productos de reconocimiento de voz adoptan dos formas: los que están diseñados para ofrecer una sencilla interfaz de voz para funciones de cómputo básicas, y aquellos que puede aceptar un dictado vocal e insertar el texto hablado en una aplicación como un procesador de palabras"[5]

[5] http://www.mailxmail.com/curso/informatica/hardwareaudio/capitulo20.htm

La tarjeta para el reconocimiento de voz es excelente ya que permite captar señales de audio claras, además de entender una serie de comandos limitados de voz de varias personas.

2.3.1.1. Características

Un chip de sonido en la tarjeta madre o un adaptador de audio en una tarjeta de expansión PCI o ISA que se instala dentro de una ranura de bus en la computadora.

2.3.1.2. Funcionamiento

La tarjeta de sonido en este caso permite el ingreso de información audible al computador, esta información es vocalizada o suministrada por el usuario la cual entra analógicamente a la tarjeta y esta permite cambiarla a digital facilitando su uso y manipulación posteriormente.

2.3.1.3. Configuración

Sin importar el programa que elija, algunos de los prerrequisitos para el éxito comprenden los siguientes:

- **Una tarjeta de sonido de alta calidad**; La cual permitirá obtener una buena definición en cuanto a sonido se refiere, ya que esta tarjeta es recomendada para trabajar con reconocimiento de voz.
- **Una habitación silenciosa** con pocos cambios en el sonido ambiental.

- **Un micrófono** diseñado especialmente para reconocimiento (el software normalmente incluye uno de bajo costo, pero existen otros mejores). Usar un micrófono de calidad deficiente o una tarjeta de sonido inadecuada ocasionará problemas, incluso al mejor programa de dictado. Para mejores resultados, use un micrófono con características de cancelación de ruido o direccional, y elija un modelo que esté en la lista aprobada por el distribuidor del software de reconocimiento de voz. Por lo regular esos micrófonos emplean una diadema para el ajuste preciso del ángulo y distancia a la boca del micrófono, lo que es un elemento crucial en un buen reconocimiento de voz.

- Llevar a cabo, si se ofrece, la **configuración de sonido** y la capacitación de usuario, para adaptar el software a su voz y vocabulario.

- **Practicar en la pronunciación** de las palabras difíciles de distinguir. Utilizar la **versión más avanzada del software**.

2.4. Inteligencia Artificial

La Inteligencia Artificial comenzó como el resultado de la investigación en psicología cognitiva y lógica matemática. Se ha enfocado sobre la explicación del trabajo mental y construcción de algoritmos de solución a problemas de propósito general. Punto de vista que favorece la abstracción y la generalidad.

La Inteligencia Artificial es una combinación de la ciencia del computador, fisiología y filosofía, tan general y amplio como eso, es que reúne varios campos (robótica, sistemas expertos, por ejemplo), todos los cuales tienen en común la creación de máquinas que pueden "pensar".

2.4.1. Funcionamiento de un Sistema de Inteligencia Artificial

El sistema inteligente tiene la capacidad de aprender por si solo del medio exterior ayudando de mejor manera al ser humano a alcanzar sus objetivos.

Hay que tomar en cuenta que el sistema inteligente únicamente tiene un objetivo temporal que deriva de uno principal, ya que posee solo unos pocos sentidos que únicamente captan luz y sonido de un objeto mas no el objeto como tal.

El sistema luego almacena estas impresiones sensoriales como conceptos elementales. Los conceptos son una forma material de almacenar información. Trabajando con conceptos, el sistema crea nuevos conceptos y almacena la relación que tienen éstos con otros conceptos totales, parciales, abstractos y concretos.

2.4.2. Reglas

Las reglas se utilizan para examinar un conjunto de datos y solicitar nueva información hasta llegar a un diagnóstico. Hoy en día las reglas, ampliadas con el uso de marcos y con algunos conceptos propios de las redes semánticas, constituyen el método estándar para la construcción de sistemas expertos.

Un sistema basado en reglas consta de cinco elementos básicos:

2.4.2.1. Motor de inferencia o intérprete de reglas:

Éste motor de inferencia permite al sistema inteligente tener una razonamiento propio acerca de un problema específico, es así que ARAYA define que "El motor de inferencia es el algoritmo o conjunto de algoritmos que implementan las reglas de razonamiento por las cuales el experto llega a sus conclusiones sobre un problema determinado". [HESS ARAYA, 2001][6]

[6] HESS ARAYA, "Inteligencia artificial y derecho", en Revista Electrónica de Derecho e Informática numero 39, octubre del 2001.

Siendo por lo tanto el elemento central del sistema, que se encarga de coordinar la información y de enviar los resultados de la inferencia al lugar oportuno.

2.4.2.2. Base de conocimiento:

Contiene las reglas y, a veces, también algunas afirmaciones iniciales. Es específica del dominio en que el sistema puede considerarse experto.

2.4.2.3. Base de afirmaciones o memoria de trabajo:

Contiene las afirmaciones iniciales, las que se extraen de la base de datos, las que el usuario introduce como datos y todas las que el motor de inferencia infiere a partir de las anteriores aplicando las reglas.

2.4.2.4. Interfaz de usuario:

Se encarga de solicitar al usuario la información necesaria y de mostrarle los resultados de la inferencia.

2.4.2.5. Base de datos:

Es necesaria para tener acceso eficiente a toda la información relativa a los casos anteriores.

2.4.3. Algoritmos de Búsqueda

Un algoritmo de búsqueda está diseñado para localizar un elemento concreto dentro de una estructura de datos. Soluciona un problema booleano de existencia o no de un elemento determinado en un conjunto finito de elementos, es decir al finalizar el algoritmo este debe decir si el elemento en cuestión existe o no en ese conjunto (si pertenece o no a él), además, en caso de existir, el algoritmo podría proporcionar la localización del elemento dentro del conjunto.

Este problema puede reducirse a devolver la existencia de un número en un vector.

2.4.4. Reconocimiento de Voz

El reconocimiento de voz adopta dos formas: que está diseñado para ofrecer una sencilla interfaz de voz para funciones de cómputo básicas, y además que puede aceptar un dictado vocal e insertar el texto hablado en una aplicación como un procesador de palabras.

Una simple congestión hace variar el tono de voz, aunque sea levemente, y el sistema no es capaz de decidir si el acceso ha de ser autorizado o no; incluso el estado anímico de una persona varía su timbre.

Los factores que dificultan el reconocimiento de voz son:

- Ambiente/canal
- Conjunto de usuarios
- Modo de elocución
- Vocabulario
- Complejidad de la tarea

2.4.4.1. Diseño de un Sistema de Reconocimiento de Voz

El hardware de audio para PC actual podría tomar una de las siguientes formas:

- Un chip de sonido en la tarjeta madre o un adaptador de audio en una tarjeta de expansión PCI o ISA que se instala dentro de una ranura de bus en la computadora.
- Un micrófono
- Parlantes

2.4.4.2. Aprendizaje

El reconocimiento de la voz requiere un breve período de "entrenamiento" de la computadora: una vez cargado el software y conectado el micrófono, hay que dejar que el programa guíe al usuario, pidiéndole, entre otras cosas, leer un texto en voz alta ante el micrófono. De este modo, el programa se adapta

a cada voz. Una vez finalizado el entrenamiento, podrá usted dictar claramente, pero en tono y ritmo naturales, sin hacer pausas entre palabras como exigían los primeros programas por el estilo.

Al hablar frente al micrófono, la voz se envía a la computadora como una señal analógica. El Software de reconocimiento de voz configurado en el idioma que se está utilizando convierte la señal analógica en una digital y la procesa para ecualizar el volumen, el tono, la velocidad y otras variantes. Un modelo acústico elige las palabras que se ajustan mejor al sonido de lo que ingresado. Un modelo de lenguaje utiliza el contexto de las palabras y la información acerca de la frecuencia para encontrar opciones probables.

Luego, un mecanismo de voz combina la información acústica y del lenguaje para adivinar las palabras que se ha pronunciado. El programa escribe en pantalla su mejor opción.

Para finalizar existe una etapa de retroalimentación cuando se corrige una palabra que no se reconoció en forma adecuada, la computadora modifica los modelos acústicos y de lenguaje para lograr mayor precisión la próxima vez.

2.4.4.3. Modelo del lenguaje

El lenguaje natural es un fenómeno muy complejo, pero ha sido sobradamente demostrado que las expresiones del lenguaje humano están organizadas a través de un conjunto de reglas.

Todas nuestras expresiones tienen una clara organización: las palabras en una oración se asocian para describir objetos y acciones, posiblemente complejas. El objetivo de un analizador sintáctico es precisamente descubrir estas asociaciones entre palabras, lo que se conoce como estructura sintáctica.

Un analizador sintáctico es un programa que toma como entrada una oración y trata de descubrir la estructura sintáctica que explica las relaciones entre las palabras de esa oración.

Los analizadores buscan la estructura correcta dentro de un conjunto de análisis posibles, este conjunto esta usualmente definido por una gramática. El modelo de lenguaje en el cual se basa el analizador sintáctico decide cuáles son los componentes sintácticos de las oraciones y como éstos están relacionados

2.4.4.4. Hardware – Características

Componentes

La figura siguiente **Figura 2** se muestra un diagrama simplificado de los componentes típicos de una tarjeta de sonido. En él se indica cuál es la información que viaja por cada enlace.

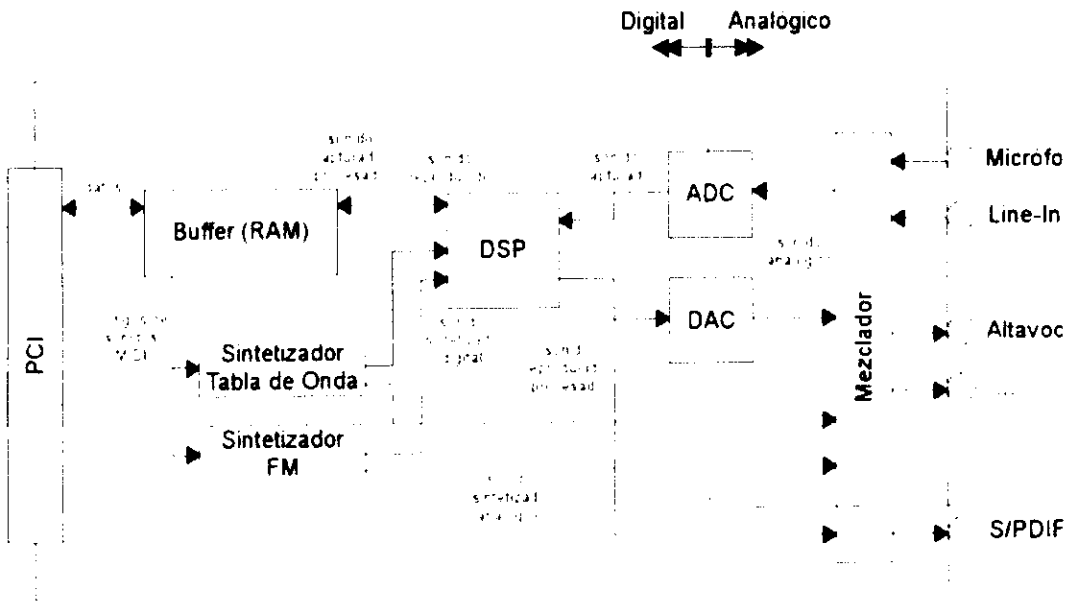


Figura 2 Diagrama de los componente de una tarjeta de sonido

Interfaz con placa madre

Sirve para transmitir información entre la tarjeta y el computador. Puede ser de tipo PCI, ISA, USB, etc.

Buffer

La función del buffer es almacenar temporalmente los datos que viajan entre la máquina y la tarjeta, lo cual permite absorber pequeños desajustes en la velocidad de transmisión. Por ejemplo, si la CPU no envía un dato a tiempo, la tarjeta puede seguir reproduciendo lo que tiene en el buffer; si lo datos llegan demasiado rápido, se van guardando. Lo mismo pasa en sentido inverso.

Muchas computadoras realizan la transmisión por DMA. Esto permite transportar los datos entre la tarjeta y la memoria directamente, sin la intervención de la CPU, lo cual le ahorra trabajo.

DSP

Procesador de señal digital. Es un pequeño microprocesador que efectúa cálculos y tratamientos sobre la señal de sonido, liberando así a la CPU de ese trabajo. Entre las tareas que realiza se incluye compresión (en la grabación) y descompresión (en la reproducción) de la señal digital. También puede introducir efectos acústicos tales como coros, reverberación, etc., a base de algoritmos.

Los DSP suelen disponer de múltiples canales para procesar distintos flujos de señal en paralelo. También pueden ser full-dúplex, lo que les permite manipular datos en ambos sentidos simultáneamente.

ADC

Conversor analógico-digital. Se encarga de transformar la señal de sonido analógica en su equivalente digital. Esto se lleva a cabo mediante tres fases: muestreo, cuantificación y codificación. Como resultado se obtiene una secuencia de valores binarios que representan el nivel de tensión en un momento concreto.

El número de bits por muestra es fijo, y suele ser 16. La frecuencia de muestreo se puede controlar desde el PC, y normalmente es una fracción de 44.1kHz.

DAC

Conversor digital-analógico. Su misión es reconstruir una señal analógica a partir de su versión digital. Para ello el circuito genera un nivel de tensión de salida de acuerdo con los valores que recibe, y lo mantiene hasta que llega el siguiente. En consecuencia se produce una señal escalonada, pero con la suficiente frecuencia de muestreo puede reproducir fielmente la original.

Sintetizador FM (modulación de frecuencia)

La síntesis por modulación de frecuencias implementa uno de los métodos de sintetizar sonido a partir de información simbólica (MIDI). Su funcionamiento consiste en variar la frecuencia de una onda portadora sinusoidal en función de una onda moduladora. Con esto se pueden conseguir formas de onda complejas con múltiples armónicos, que son lo que define el timbre. El tono y volumen del sonido deseado los determinan la frecuencia fundamental y la amplitud de la onda.

Los primeros sintetizadores FM generaban una señal analógica. Sin embargo, posteriormente se han desarrollado versiones que trabajan digitalmente. Esto da más flexibilidad y por tanto más expresividad a la generación de ondas, a la vez que permite someter la señal a tratamiento digital.

Sintetizador por Tabla de Ondas

La síntesis mediante tabla de ondas es un método alternativo al FM. En vez de generar sonido de la nada, utiliza muestras grabadas de los sonidos de instrumentos reales. Estas muestras están almacenadas en formato digital en una memoria ROM incorporada, aunque también pueden estar en memoria principal y ser modificables. El sintetizador busca en la tabla el sonido que más se ajusta al requerido en cada momento. Antes de enviarlo

realiza algunos ajustes sobre la muestra elegida, como modificar el volumen, prolongar su duración mediante un bucle, o alterar su tono a base de aumentar o reducir la velocidad de reproducción.

Este componente puede tener una salida analógica o digital, aunque es preferible la segunda. En general el sonido resultante es de mayor calidad que el de la síntesis FM.

Alternativamente, este proceso puede ser llevado a cabo enteramente por software, ejecutado por la CPU con muestras almacenadas en disco y un algoritmo apropiado (códecs de audio). Esta técnica es muy utilizada porque permite abaratar el coste de la tarjeta.

Mezclador

El mezclador tiene como finalidad recibir múltiples entradas, combinarlas adecuadamente, y encaminarlas hacia las salidas. Para ello puede mezclar varias señales (por ejemplo, sacar por el altavoz sonido reproducido y sintetizado) o seleccionar alguna de ellas (tomar como entrada el micrófono ignorando el Line-In). Este comportamiento se puede configurar por software.

Tanto las entradas como las salidas pueden proceder de la tarjeta o del exterior. El mezclador suele trabajar con señales analógicas, aunque también puede manejar digitales (S/PDIF).

Conectores

Son los elementos físicos en los que deben conectarse los dispositivos externos, los cuales pueden ser de entrada o de salida.

Casi todas las tarjetas de sonido se han adaptado al estándar PC 99 de Microsoft que consiste en asignarle un color a cada conector externo, como se muestra en la **Figura 3**:

Color	Función
Rosa	Entrada analógica para micrófono.
Azul	Entrada analógica "Line-In"
Verde	Salida analógica para la señal estéreo principal (altavoces frontales).
Negro	Salida analógica para altavoces traseros.
Plateado	Salida analógica para altavoces laterales.
Naranja	Salida Digital SPDIF (que algunas veces es utilizado como salida analógica para altavoces centrales).

Figura 3: Esquema de colores en la tarjeta de Sonido Sound blaster

Los conectores más utilizados para las tarjetas de sonido a nivel de usuario son los mini-jack al ser los más económicos. Con los conectores RCA se consigue mayor calidad ya que utilizan dos canales independientes, el rojo y el blanco, uno para el canal derecho y otro para el izquierdo.

A nivel profesional se utilizan las entradas y salidas S/PDIF, también llamadas salidas ópticas digitales, que trabajan directamente con sonido digital eliminando las pérdidas de calidad en las conversiones.
Para poder trabajar con dispositivos MIDI se necesita la entrada y salida MIDI.

2.4.4.5. Software – Características

El reconocimiento de voz se acerca más a la corriente principal y encuentra su mercado en las que se emplean vocabularios especializados, pero sin desechar aún el teclado y el mouse. Estos programas están diseñados para complementar los medios de captura tradicionales, no para reemplazarlos.
Además pueden ofrecer un gran incremento en la productividad de los usuarios con habilidades limitadas para mecanografiar. Pero para aquellos que tienen una discapacidad, y que sufren daños constantes por el estrés o simplemente siempre tienen las manos ocupadas estos productos son magníficas.

Según la revista PC MAGAZINE se pusieron a prueba, las actualizaciones más recientes de los cuatro productos de reconocimiento de voz: **Dragon Naturallly Speaking Preferred 4.0, Free Speech 2000 de Philips, L & H Voice Xpress Profesional Versión 4.0 y Via Voice Pro Millenium Edition de IBM.** Todos los fabricantes refinaron sus mecanismos de voz y mejoraron la facilidad de uso e incrementaron el soporte destinado a comandos y controles, así como para el Natural Language Command (NLC) de modo que funcionen en una extensa variedad de aplicaciones, como Microsoft Excel y Outlook.[7]

2.4.4.6. Usos y Aplicaciones

Los usos y aplicaciones que se le dan a cada software de reconocimiento de voz depende de las características de cada uno, por tal motivo en los laboratorios de la PC MAGAZINE pusieron a prueba las actualizaciones mas recientes teniendo como resultado final las siguientes características a resaltar.

Dragon Naturally Speaking Preferred 4.0

Este software es uno de los mejores para el reconocimiento de palabras y comandos por voz. Este producto ha optimizado la utilización de más librerías para una buena recepción auditiva. Sólo tiene la desventaja de un

[7] http://www.itq.edu.mx/vidatec/espacio/aisc/ARTICULOS/ASCC/articuloASCC.htm

soporte limitado para el lenguaje natural en aplicaciones distintas de Microsoft Word.

Requerimientos:

Pentium MMX/200 o equivalente; 48 MB en RAM; 200MB de espacio en disco duro; tarjeta de sonido de 16 bits compatible con Creative Labs Sound Blaster; Microsoft Windows 95 , 98 o NT.

Dragon Naturally Preferred 4.0, marca el estándar para la facilidad de uso, desde la instalación directa hasta el formato sencillo y la navegación en el Web activada por la voz. Ahora Dragon Naturally Speaking 9.0 ya no necesita que le entrenen ya que directamente se puede adaptar a cualquier frecuencia de voz.

FreeSpeech 2000

Requerimientos:

Pentium/MMX 166 o equivalente; 48 MB en RAM; 100MB de espacio en disco duro; tarjeta de sonido compatible con Sound Blaster; Microsoft Windows 95,98 o NT con SP3.

El modo de dictado tiene una característica EasyEdit, que destaca las palabras mientras reproduce el dictado, y la característica EasyCorrect del programa le permite interrumpir la reproducción y reemplazar la palabra

subrayada con la opción seleccionada de una lista de alternativas. El botón de SpeechMike facilita el proceso, excepto que, cuando lo suelta, el micrófono se apaga y debe pasar en forma manual al modo de Dictado, lo que reduce los beneficios del botón. Para deletrear palabras, debe cambiar al modo Deletrear, y si desea interrumpir la grabación de la voz, recurra al modo Inactivo. Utilizar cuadros de diálogo que requieren una combinación de navegación y dictado puede ser difícil.

L & H Voice Xpress Professional, Versión 4

Requerimientos:

Pentium II; 48 MB en RAM con Microsoft Windows 95 o 98, o 64 MB con Windows NT; 200 MB de espacio en disco duro; tarjeta de sonido de 16 bits compatible con Sound Blaster o micrófono USB.

L & H Voice Xpress Professional ofrece las características básicas que se espera de un programa de reconocimiento de voz: es posible dictar en cualquier aplicación, crear macros de voz y controlar las aplicaciones con comandos de voz. El punto fuerte del programa son sus comandos intuitivos con lenguaje natural para Microsoft Word, Excel y el sencillo procesador de palabras Voice Xpress. Pero sus características de corrección son limitadas; por ejemplo, carecen de reproducción de audio.

Via Voice Pro Millennium Edition

Requerimientos:

Pentium/233 o equivalente; 48 MB en RAM con Microsoft Windows 95 o 98, o 64 MB con Windows NT; 100 MB de espacio en disco duro; tarjeta de sonido de 16 bits.

Via Voice Pro Millenium Edition de IBM, ofrece una amplia variedad de funcionalidad para aplicaciones de dictado y de control. Los novatos empezarán a trabajar con rapidez, mientras que los usuarios de poder apreciarán las capacidades avanzadas de macros del programa. No obstante, los evaluadores encontraron que algunas de las características de edición de Via Voice son complicadas, un factor determinante en su facilidad de uso en general. La instalación y el manejo inicial transcurrieron sin problemas gracias a las pantallas asistente bien diseñado y un personaje animado que explica cómo empezar. Los usuarios podrán empezar a dictar de 20 a 30 minutos después de la instalación del software, lo que incluye de 10 a 15 minutos de lectura del texto del registro.

2.4.5. Procesamientos de Lenguajes Naturales

El **Procesamiento de Lenguajes Naturales**, abreviado **PLN**, o **NLP** del idioma inglés Natural Language Processing, es una subdisciplina de la Inteligencia Artificial y la rama ingenieril de la lingüística computacional. El

PLN se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas o entre personas y máquinas por medio de lenguajes naturales. El PLN no trata de la comunicación por medio de lenguajes naturales de una forma abstracta, sino de diseñar mecanismos para comunicarse que sean eficaces computacionalmente, que se puedan realizar por medio de programas que ejecuten o simulen la comunicación. Los modelos aplicados se enfocan no sólo a la comprensión del lenguaje de por sí, sino a aspectos generales cognitivos humanos y a la organización de la memoria. El lenguaje natural sirve sólo de medio para estudiar estos fenómenos.

2.4.5.1. Dificultades en los PLN

2.4.5.1.1. Ambigüedad

El lenguaje natural es inherentemente ambiguo a diferentes niveles:

A nivel léxico, una misma palabra puede tener varios significados, y la selección del apropiado se debe deducir a partir del contexto oracional o conocimiento básico. Muchas investigaciones en el campo del procesamiento de lenguajes naturales han estudiado métodos de resolver las ambigüedades léxicas mediante diccionarios, gramáticas, bases de conocimiento y correlaciones estadísticas.

A nivel referencial, la resolución de anáforas y catáforas implica determinar la entidad lingüística previa o posterior a que hacen referencia.

A nivel estructural, se requiere de la semántica para desambiguar la dependencia de los sintagmas preposicionales que conducen a la construcción de distintos árboles sintácticos. Por ejemplo, en la frase "*Rompió el dibujo de un ataque de nervios*".

A nivel pragmático, una oración, a menudo, no significa lo que realmente se está diciendo. Elementos tales como la ironía tienen un papel importante en la interpretación del mensaje.

Para resolver estos tipos de ambigüedades y otros, el problema central en el PLN es la traducción de entradas en lenguaje natural a una representación interna sin ambigüedad, como árboles de análisis.

2.4.5.1.2. Detección de separación entre las palabras

En la lengua hablada no se suelen hacer pausas entre palabra y palabra. El lugar en el que se debe separar las palabras a menudo depende de cuál es la posibilidad que mantenga un sentido lógico tanto gramatical como contextual. En la lengua escrita, idiomas como el chino mandarín tampoco tienen separaciones entre las palabras.

2.4.5.1.3. Recepción imperfecta de datos

Acentos extranjeros, regionalismos o dificultades en la producción del habla, errores de mecanografiado o expresiones no gramaticales, errores en la lectura de textos.

2.4.5.1.4. Componentes

- Análisis morfológico. El análisis de las palabras para extraer raíces, rasgos flexivos, unidades léxicas compuestas y otros fenómenos.

- Análisis sintáctico. El análisis de la estructura sintáctica de la frase mediante una gramática de la lengua en cuestión.

- Análisis semántico. La extracción del significado de la frase, y la resolución de ambigüedades léxicas y estructurales.

- Análisis pragmático. El análisis del texto más allá de los límites de la frase, por ejemplo, para determinar los antecedentes referenciales de los pronombres.

- Planificación de la frase. Estructurar cada frase del texto con el fin de expresar el significado adecuado.

- Generación de la frase. La generación de la cadena lineal de palabras a partir de la estructura general de la frase, con sus correspondientes flexiones, concordancias y restantes fenómenos sintácticos y morfológicos.

CAPITULO III.

DESARROLLO DEL PROYECTO

3.1. Enfoque

Cuando se realiza un proyecto con reconocimiento de voz, es importante tener claro el propósito del mismo, esto quiere decir qué usuarios van a formar parte en la interacción con el dispositivo de reconocimiento que en este caso es el humanoide. Hay que tomar en cuenta diferentes factores, tales como fonema de cada persona, claridad en la pronunciación, volumen, acercamiento al micrófono, etc. Todos estos aspectos hacen que el reconocimiento de voz sea un proceso complejo.

La comunicación entre personas parece un proceso sencillo y como es tan común, nos olvidamos que para establecer una conversación el hombre ha tenido que evolucionar miles de años. Cosa muy diferente pasa con el reconocimiento de voz usuario – máquina, ya que cualquier sonido que haga la persona puede el programa interpretar como un ruido, mas no como un fonema, mucho menos una palabra. Por lo tanto, es necesario tomar en cuenta los componentes básicos de un sistema de reconocimiento de voz (usuario – máquina).

Convertir en texto el habla estándar del usuario es una tarea extraordinariamente difícil, dada la amplia variación en los patrones del habla humana. Por esta razón, casi todo el software de este tipo (y también algunas de las aplicaciones de comandos de voz básicos) deben ser "entrenadas" para entender la voz de un usuario en particular. Este entrenamiento consiste en leer una parte de un texto, o un artículo que por lo general el mismo software facilita para que la persona pronuncie, y se vaya adaptando la voz para cada usuario.

Debido a que el software "sabe" de antemano lo que se supone que la persona dice, puede asociar ciertas palabras con la forma en que las pronuncia (fonética).

Es así que al humanoide se va a agregar ésta facultad, de poder entender y responder lo que los usuarios requieran. Por lo tanto este proyecto se centra principalmente en brindar información al usuario en forma más directa. Permitir agregar una característica especial al humanoide, que ha sido de importancia en la Escuela de Sistemas y que ha permitido saber algo más de la Universidad, incentivando de esta manera en Colegios y Universidades lo que se puede tratar de desarrollar en la PUCESA.

El proyecto anterior se enfocó en permitir al usuario utilizar el humanoide para fines informativos, por lo tanto la persona que necesita la información debe escoger una de las cuatro opciones en el tablero de control que contienen información general de la Universidad Católica, como son "Acerca

de la PUCESA", Propósito de la Universidad, "Carreras de la PUCESA" diferentes carreras que tiene actualmente, "Entretenimiento" videos musicales, "Finalizar" que finaliza el programa que está en ejecución.

Lo que se trata en este proyecto es agregar la característica de comunicarse con las personas de una forma más interactiva y natural, sin perder las opciones que tiene el humanoide en el tablero de control.

3.2. Metodología de Desarrollo de Software

Para este proyecto se ha visto necesario trabajar con el método de prototipos, ya que permite trabajar desde su diseño hasta la integración total del hardware (robot) y software (programa) que es el objetivo final de este proyecto, facilitando hacer cambios según las pruebas con respecto a las necesidades de los usuarios.

El método de desarrollo según prototipos tiene las siguientes fases:

3.2.1. Obtención de requisitos

El programa debe vincularse con los siguientes requisitos:

- Tarjeta de sonido bien configurada

- Instalación del Agente de Windows, SDK (Software Development Kit) para la utilización del agente en el idioma en Español
- Configuración de los servomotores con el programa VSA
- Instalación y configuración del programa de reconocimiento de voz
- Micrófono en perfecto estado y configuración para realizar las pruebas

3.2.2. Diseño Global del prototipo

Las palabras ingresadas por el usuario al programa, se las realiza mediante un micrófono el cual es el complemento ideal para trabajar con el programa de reconocimiento de voz, el cual asimila estas palabras y las ingresa a un cuadro de texto (Textbox) en el formulario de programación en Visual Basic, el mismo que ha sido insertado para contener las frases y palabras que el usuario ingrese al hacer las preguntas respectivas para la interacción. Luego estas palabras van a ser comparadas en el programa conjuntamente con la base de datos, si estas palabras están en esta base, un agente de Windows que sirve para convertir en audio las palabras, va a emitir la respuesta con una voz sintetizada, siendo esta la finalización de la interacción con el usuario.

A continuación se representa en el gráfico, la secuencia de interacción del usuario con el humaoide (Máquina), siendo el micrófono el principal medio de comunicación de este Proyecto, sin dejar a un lado las opciones

principales (botones) en el tablero de control. La respuesta va a emitirse mediante los altavoces y finalmente la respuesta es receptada por el usuario.

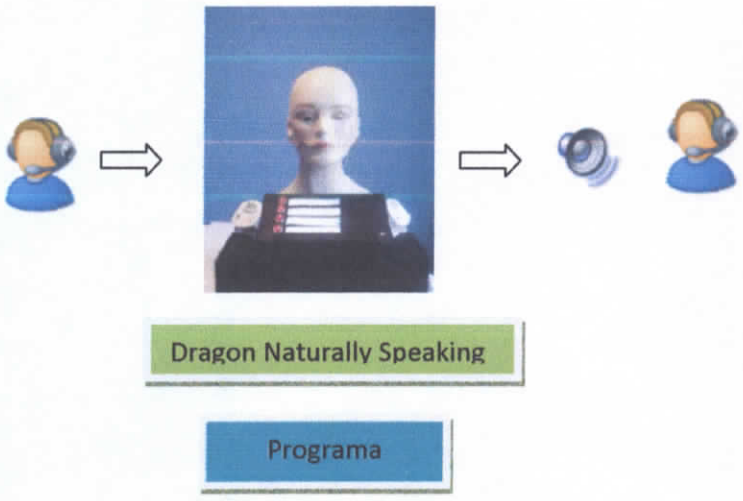


Figura 4 . Interacción Usuario - Máquina

Para que la interacción sea una mejora del proyecto anterior, se mantienen las mismas opciones, la mejora en este caso es poder contar con el reconocimiento de voz. Por lo tanto la pantalla principal y única del proyecto se muestra en la **Figura 5:**



Figura 5. Ventana Principal del Programa

Como va ser una interacción directa con el humanoide, y al mismo se lo pretende hacer más real, no se va contar con monitor, mouse ni teclado, además la pantalla principal no va ser visible, entonces el único medio de acceder al humanoide es mediante el panel de control y mediante el micrófono para el reconocimiento de voz.

3.2.3. Construcción del prototipo

Para el funcionamiento del programa en el presente proyecto debemos instalar y configurar Dragon Naturally Speaking 9.0 para que inicie junto con el programa al arrancar Windows.

Como se dijo anteriormente es necesario agregar un Cuadro de texto para que la frase reconocida por Dragon sea ingresada aquí, permitiendo que la misma mediante un proceso interno mande la respuesta.

Teniendo en cuenta estos parámetros se inicia con el desarrollo del programa de la siguiente manera:

Al programa que funcionaba anteriormente con el humanoide y que está desarrollado en Visual Basic 6.0, se le agrega los siguientes componentes:

- 1 TextBox con nombre txt_pregunta, que va a ser el objeto principal que va a contener la información de entrada.

- 1 Timer, que permitirá ejecutar el analizador léxico y además de devolver las respuesta en un tiempo determinado.

- 1 Agente de Microsoft, que se agrega dando click derecho en la caja de herramientas situada en la parte izquierda de la ventana de Visual Basic, como se presenta en la **Figura 6** y **Figura 7.**
  Figura 7

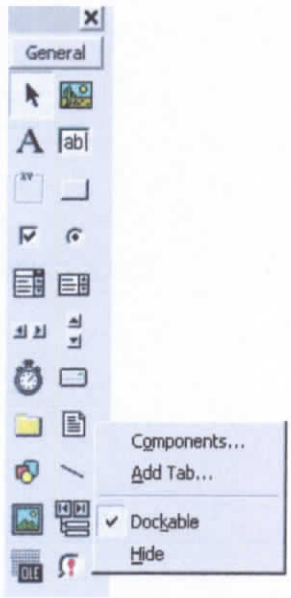
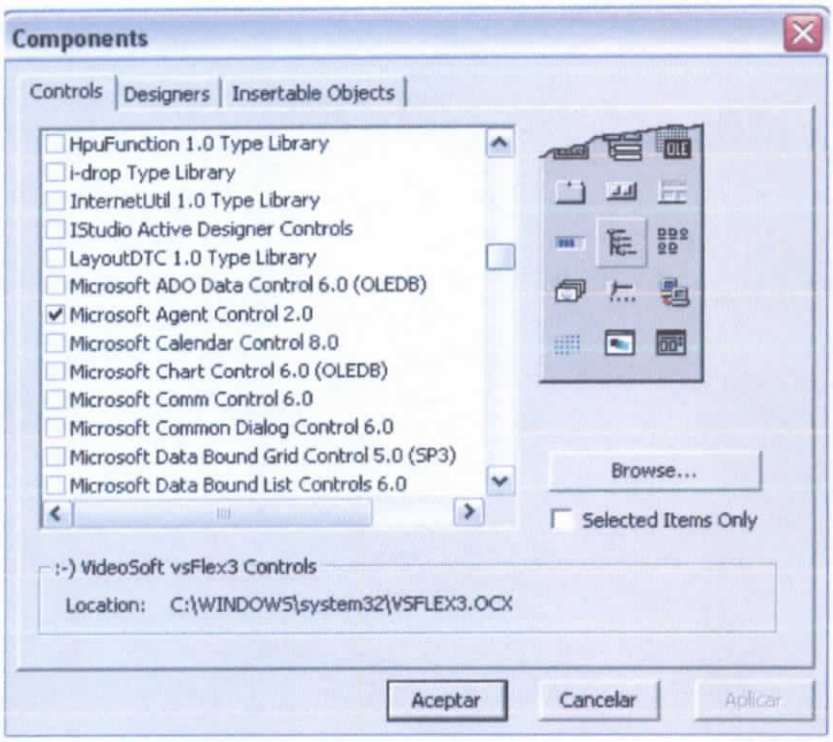


Figura 6. Caja de Herramientas

Se escoge la Opción Componente (Components dependiendo el idioma instalado), luego se aparece otra pantalla, se selecciona la opción Microsoft Agent Control 2.0.



Apareciéndose en el proyecto el siguiente el icono , luego es agregado en el formulario del programa.

Una vez que se ejecuta el programa, aparece el Agente animado que ayuda en la interacción.

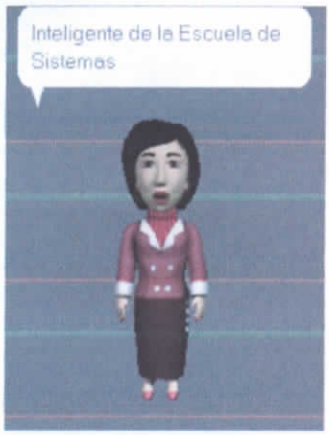
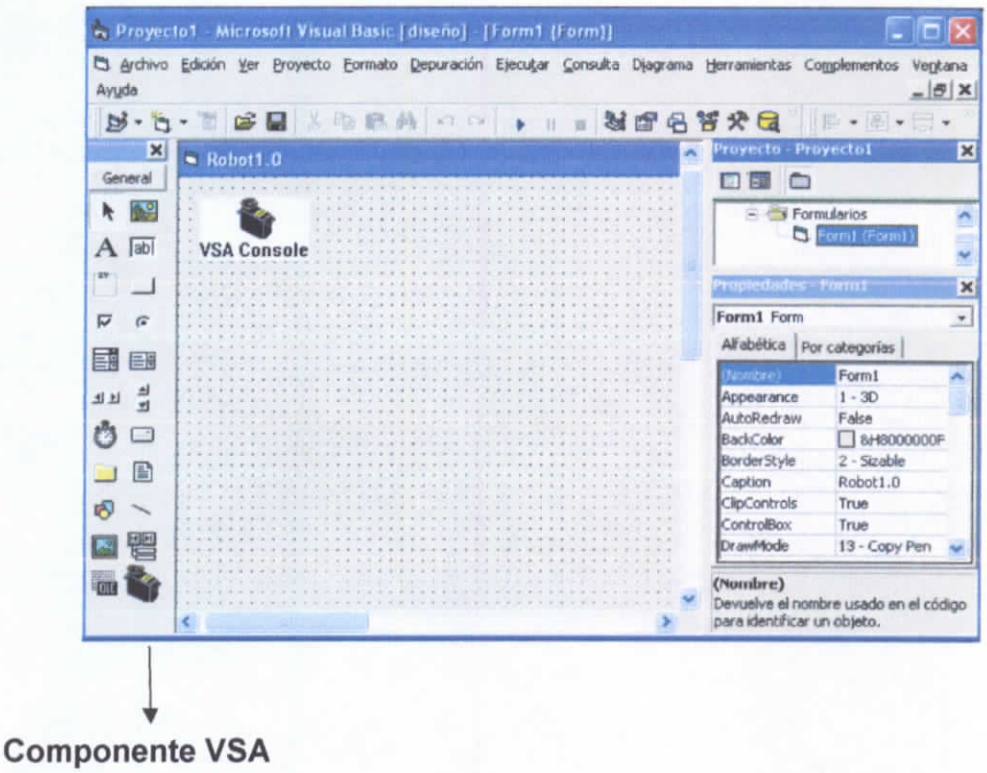


Figura 8. Microsoft Agente

- De la misma manera anterior agregar otro componente llamado Dragon Naturally Speaking Active X Controls, que permitirá configurar el encender el micrófono al inicio de la aplicación.



- Se instala la Consola OCX VSA en System32 ya que es necesaria para que Visual Basic interactué con el VSA.



Componente VSA

Figura 9. Consola VSA en el Formulario del Programa

En la figura 10 se presenta el formulario principal del proyecto en el cual se encuentran alojados todos los componentes antes mencionados, no hay que

olvidar que se trabaja a partir del proyecto de tesis anterior, únicamente con las mejoras para el reconocimiento de voz.

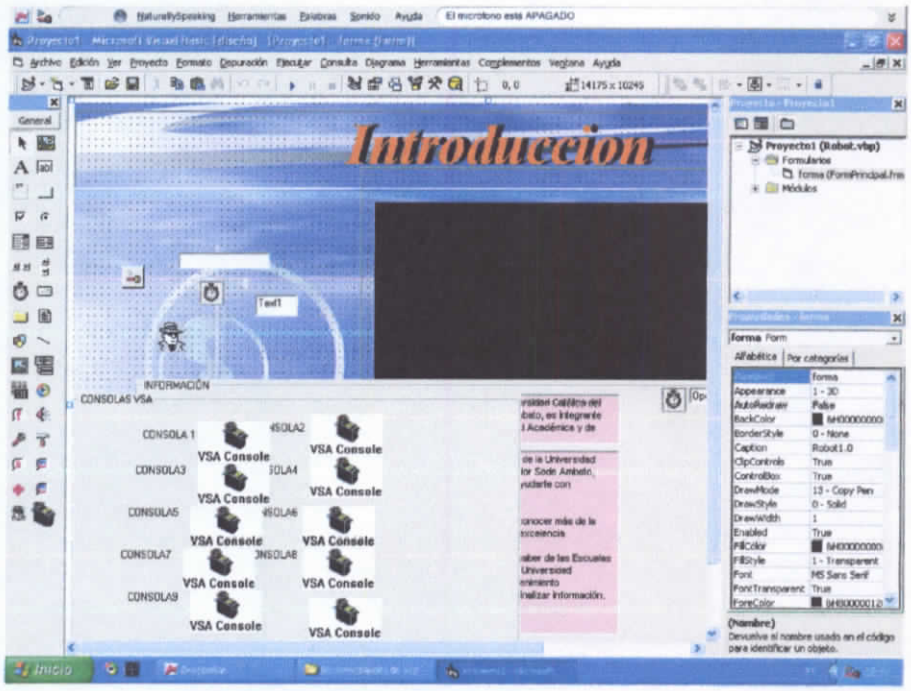


Figura 10. Formulario Principal del Programa

3.2.4. Fase de Pruebas del Prototipo

Para las pruebas hay que tener en cuenta lo siguiente:

1. Colocar el micrófono a una distancia tal que lo hablado no se distorsione o sea muy bajo.
2. Hablar claro y en tono normal para que el programa de reconozca la frase.

3. Si el programa no ha reconocido esta frase, el Agente responderá "No entiendo, repíteme por favor".
4. Luego de estos pasos, si existiera algún error en la aplicación, reiniciar la misma, o si se tarda en responder, podría ser por las siguientes causas: virus o programas que se ejecutan al inicio de Windows.

En la tabla 1 se resume el proceso de interacción con el humanoide mediante reconocimiento de voz.

Hablar claro y fuerte por el micrófono para un mejor reconocimiento 	Estas palabras vocalizadas se ingresan en un cuadro de texto. 	El analizador léxico toma esta  palabra y según la búsqueda en la base de datos emite un resultado.
Este resultado es expuesto al usuario en audio mediante el Agente de Windows. Y sincronizado con los servomotores del humanoide 		Si no encuentra la palabra ingresada se emite la respuesta "No entiendo, repíteme por favor" 

Tabla 1. Resumen de la interacción humano – máquina

3.3. Selección de Información

La información que se va a manejar en este proyecto de tesis se centra exclusivamente a la Escuela de Ingeniería de Sistemas, ya que toda la información que se consulte al humanoide va a ser con relación a la misma. Sin embargo, la estructura del software permite que en el futuro se pueda insertar nueva información.

Esta información va a ser guardada en una base de datos, la cual ha sido depurada para que sea manejada solo con información esencial de la carrera y así satisfacer las necesidades de los usuarios que necesiten conocer más acerca de la Escuela de Ingeniería de Sistemas.

A continuación se detallan los principales temas que va a contener la base de datos:

- Perfil Profesional de la Carrera
- Cantidad de Niveles para aprobar la materia
- Detalle de las materias de cada nivel
- El campo ocupacional de la carrera
- Títulos que otorga la carrera
- Nombre de los docentes que trabajan en la carrera
- Pensum Académico
- Beneficios de la Universidad

- Información acerca de eventos que organiza la escuela
- Entretenimiento

3.4. Procesamiento de la Información

Una vez almacenada la información en la base de datos es necesario saber cómo realiza la búsqueda de la mejor información para satisfacción de los usuarios que necesitan saber datos puntuales como generales. Para tener como resultado la información puntual del tema a consultar, el proceso de reconocimiento de voz es únicamente el inicial, ya que una vez identificada la palabra vocalizada por el usuario, el programa compara la frase almacenada en una base de datos que regresa con la información que corresponde a la palabra clave, una vez ubicada la respuesta el programa ejecuta la respuesta en forma vocalizada por el robot, brindado la información requerida por el usuario.

Por lo tanto la secuencia de operación usuario máquina se muestra en la Figura 11.

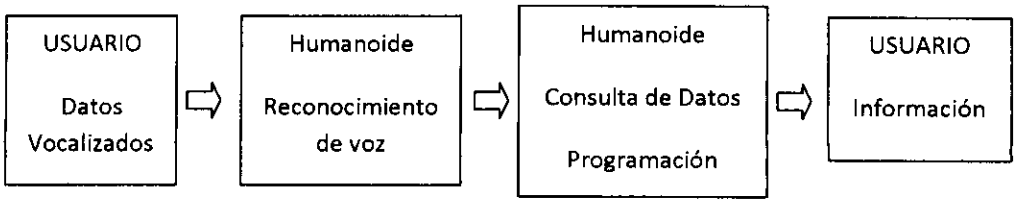
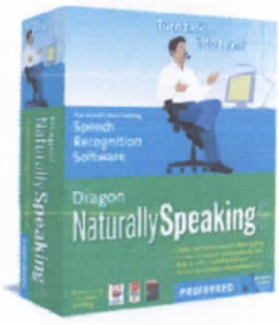


Figura 11. Secuencia de Interacción Maquina – Usuario

3.5. Software utilizado en la Implementación

Este proyecto requiere tres herramientas importantes y necesarias para este proyecto, como es el Dragon Naturally Speaking que permite dictar y ver al instante las palabras ya sea en un documento o correo electrónico. La versión con la que se trabaja en este proyecto es especial ya que no necesita un entrenamiento previo para comenzar a utilizar este programa.



Otra herramienta esencial para el funcionamiento del programa es Visual Basic, que es un software para desarrollo de aplicaciones que en su versión 6.0 permite vincular de manera óptima con el hardware y además con las API (Interfaz de Programación de Aplicaciones) de Windows necesarias para este programa.

Por ultimo el Software que almacena toda la información del proyecto como es Microsoft Access, una herramienta básica pero muy eficiente al trabajar con aplicaciones Microsoft como es Visual Basic.

3.5.1. Dragon Naturally Speaking

La información ingresada por el usuario, es de forma oral por lo que es necesario valerse de un programa para el reconocimiento de voz como es el Dragon Naturally Speaking en su versión 9, permitiendo así ingresar estos datos vocalizados por el usuario. Esta versión se ha escogido en este proyecto de tesis ya que desde su instalación es un software fácil de utilizar. Lo importante para destacar de Dragon Naturally Speaking 9 Preferred, es que no necesita de un entrenamiento inicial para comenzar a trabajar en él, por lo tanto es el indicado para utilizar en el reconocimiento de voz para varios usuarios, facilitando en este proyecto la interacción con el humanoide, además se adapta fácilmente a los programas de edición de texto de Microsoft y también la navegación web que es activada por voz.

Existe actualmente otro software para el reconocimiento de voz como es el IBM Via Voice Pro 8, que es también uno de los mejores que existen actualmente y para que este programa funcione con precisión, el mismo debe adaptarse con la voz de una persona específica, cosa que no ayudaría de mucho si se quiere trabajar con multi-usuarios, y si quieren trabajar varias personas con este programa deben entrenar una sesión de usuario con cada persona.

Para el reconocimiento de voz se puede trabajar a nivel de hardware, con una tarjeta de reconocimiento de voz, que es un circuito receptor de señales analógicas a la cual se la conecta directamente un micrófono. Es muy buena

en la modulación de las señales de entrada, esto quiere decir que reconoce la voz de casi todas las personas, pero la gran desventaja de utilizar esta tarjeta es que por ser hardware no puede guardar mucha información es decir no puede almacenar muchos comandos, reduciendo de esta manera los datos a almacenar, siendo esta razón un punto en contra con el objetivo del proyecto en brindar información de la carrera con opciones a aumentar en el futuro.

3.5.2. Microsoft Visual Basic 6.0

En este proyecto, se utiliza además Visual Basic 6.0 para la programación del analizador léxico que selecciona los datos fundamentales para la interacción de usuarios con la máquina, además este lenguaje se ha utilizado en el proyecto anterior y como se trata de hacer una mejora, se continúa desarrollando con el mismo software de programación. Otro aspecto importante de este software es que es idóneo para trabajar directamente con hardware, y también funciona con la mayoría de programas en Microsoft Windows.

3.5.3. Microsoft Access

El respaldo de la información va a ser guardada en la base de datos de Microsoft Access, que es una herramienta sencilla de utilizar y que la conexión con Visual Basic 6.0 no trae complicaciones.

3.6. Tipos de Usuarios

El reto más importante del proyecto es el reconocimiento de voz de cualquier usuario que desee conocer información acerca de la carrera de Ingeniería de Sistemas, por lo tanto hay que tomar en cuenta lo siguiente:

- El tono de voz
- Semántica
- Fonética

Estas son características muy importantes para el mejor reconocimiento de la voz. La tarea de un sistema de reconocimiento de voz es muy compleja ya que debe captar señales acústicas, fonéticas, fonológicas y posiblemente léxicas, con los correspondientes procedimientos interpretativos para cada persona.

En este caso el humanoide debe reconocer todas las voces que interactúen con el mismo, como estudiantes, profesores, usuarios en general, dentro o fuera de la Universidad. Además, no hay que olvidar que las personas poseen voces agudas y graves y se diferencian unas de otras, tarea que el sistema de reconocimiento de voz debe controlar.

3.7. Análisis de la Estructura de los Datos

Como se va a trabajar con información acerca de la Escuela de Sistemas, es necesario utilizar una base de datos que permita guardar aquella información útil para el usuario. Es así que se va a trabajar con palabras claves ingresadas por el usuario en forma oral, las mismas que van a ser reconocidas por el sistema de reconocimiento de voz, analizadas y comparadas desde Visual Basic mediante código SQL a la Base de Datos de Access, la cual nos enviará la respuesta siempre y cuando sea la que corresponda con la respectiva clave.

La estructura de la Base de Datos es la siguiente:

- Dos tablas (comunes y conocimiento)
- La tabla comunes va almacenar todas la palabras que se encuentran comúnmente en la oración, como por ejemplo: artículos (el, la, los, las), adverbios (cómo, cuándo, dónde, porqué), contracción (del, al).
- La tabla conocimiento va a contener cuatro campos esenciales, los cuales son: "claves" que almacenan las palabras ingresadas por el usuario y serán las posibles preguntas de los usuarios al querer saber mas acerca de la carrera, el campo "respuesta", que obviamente va a almacenar la respectiva respuesta del campo claves, y el campo "tipo" que va guardar la información de que tipo de frase es, si es larga, mediana o corta dependiendo de esto y el número de palabras en la frase se le asigna un número que especifica a qué archivo del VSA activar.

Código (Primary Key)	Nombre
NUMERO	TEXTO
...	...

Tabla 2: Estructura de la Base de Datos COMUNES

Código (Primary Key)	Claves	Respuesta	Tipo
NUMERO	TEXTO	TEXTO	NUMERO
...	...	...	...

Tabla 3: Estructura de la Base de Datos CONOCIMIENTO

3.8. Comandos de Voz

Las preguntas del usuario que son ingresadas por el micrófono al humanoide van a ser procesadas por un analizador léxico, el cual comparará si la sentencia ingresada va a ser una orden o comando de voz. Se da el término de comando de voz ya que cualquier dato ingresado en forma oral al humanoide va a tener su respectiva respuesta o acción, que en este caso

resulta ser la respuesta almacenada en la base pero la misma va a ser convertida a voz.

Para diferenciar el texto ingresado de un comando de voz, lo primero que hay que tomar en cuenta es lo siguiente:

- Todo texto ingresado va a contener pronombres interrogativos, artículos, preposiciones, etc.; los cuales no van a ser tomados en cuenta para la comparación en la base ya que estas palabras son comunes en toda frase, es así que se toma en cuenta las palabras que se relacionan con la carrera de Ingeniería de Sistemas, como por ejemplo se explica a continuación:

"Cuantos niveles tiene la carrera de Ingeniería de Sistemas". Aquí se encuentran tres frases importantes una de ellas es "niveles", que si es ingresada sola no va a dar una respuesta si no es con la palabra "carrera" o "Ingeniería de Sistemas", ya que de esta forma se sabrían específicamente de que se esta hablando.

Como este proyecto almacena datos de la carrera, todas las frases relacionadas a ésta, van a ser tomadas en cuenta, como por ejemplo: niveles, pensum, profesores, etc. Permitiendo de esta forma respuestas específicas y concisas.

3.9. Analizador Léxico

El analizador léxico es el encargado de la extracción de **tokens** (palabras, signos de puntuación, constantes, operadores, que se identifican como relevantes dentro de una palabra) y la eliminación de las palabras no relevantes (comunes) y finalmente la extracción de palabras clave (keywords). Este analizador se iniciará siempre que haya cambio en la entrada de texto del programa a medida que el usuario vaya preguntando al humanoide. Una vez ingresada una frase en este cuadro de texto, el programa mediante un Timer (evento que ejecuta cierta acción en un determinado tiempo), ejecutará siempre que haya un cambio en el texto ingresado, por lo tanto cada frase ingresada va a ser reconocida por el analizador léxico, teniendo siempre una respuesta después ingresada la frase, luego de realizar esta comparación, la frase se borra automáticamente para dar lugar a la nueva frase que va a ser ingresada por el usuario mediante el programa de reconocimiento de voz como es el Dragon Naturally Speaking.

El código del analizador léxico de este proyecto que está desarrollado en Visual Basic 6.0 tiene la siguiente estructura:

```
Private Sub Timer2_Timer()
' PROCESO QUE EXTRAE TOKENS
Dim frase As String
Dim posicion As Integer
Dim palabra As String
Dim derecha As String
Dim longitud As Integer
Dim claves As String
Dim basura As Recordset
Dim consulta As Recordset
Dim existe As Integer
Dim hay As Integer

claves = " "
frase = txtpregunta.Text & " "
longitud = Len(frase)
Do While longitud > 0
  posicion = InStr(1, frase, " ", vbTextCompare)
  palabra = Left(frase, posicion - 1)


  ' consulto si es basura
  Set basura = Conn.Execute("Select * from comunes Where nombre = '" & palabra & "'  ")
  existe = 0
  Do While Not basura.EOF
   existe = 1
   basura.MoveNext
  Loop
```

```
  If existe = 0 Then
    claves = claves & " " & palabra
    claves = Trim(claves)
  End If

  derecha = Right(frase, longitud - posicion)
  frase = derecha
  longitud = Len(frase)

Loop

Set consulta = Conn.Execute("Select * from conocimiento Where Claves like '" & claves & "'
")
 hay = 0

  Do While Not consulta.EOF
  hay = 1
  dr.Speak consulta!respuesta
  txtpregunta.Text = ""
  Timer2.Enabled = False


  If consulta!tipo = 1 Then
   forma.Console11.Create
    forma.Console11.play 1, False
  ElseIf consulta!tipo = 5 Or consulta!tipo = 4 Then
    forma.Console11.Create
    forma.Console11.play 1, False
  End If
```

```
    consulta.MoveNext
Loop

If hay = 0 Then
   If Len(Trim(txtpregunta.Text)) > 0 Then
     dr.Speak "No te endiendo, !!!!"
     Timer2.Enabled = False
     End If
End If

Do While Not consulta.EOF
    hay = 1
    Rem idc = consulta!codigo
    Timer2.Enabled = False
Loop


txtpregunta.Text = " "
End Sub
```

Tabla 4: Código Fuente de la extracción de tokens (palabras relevantes)

3.10. Herramientas de Interacción y Voz

Mediante un Agente de Windows que es un ayudante animado y programado para ayudar a los usuarios a interactuar con los programas, como en Microsoft Word que facilitan la búsqueda de información en el

contenido de la ayuda, siendo una manera de entretenimiento mientras los usuarios realizan sus trabajos. Lo interesante de los mismos es que pueden hablar y además permite pasar las palabras almacenadas en texto a audio esto lo puede hacer posible la librería Text To Speech que permite a Visual Basic ejecutar el habla convirtiéndolas en audio. Se puede tener diferentes voces y en diferentes lenguajes por medio del paquete de desarrollo de Windows como es el SDK (Sotware Development Kit) SAPI (Speech API) que son funciones y procedimientos que ejecutan el habla de los personajes. Existen diferentes tipos de personajes, como son: el famoso Merlín, el Genio, Clipito etc.



Figura 12: Merlín



Figura 13: Genio

Estos personajes van hablar el texto almacenado en la base de datos, por lo tanto sólo se va a escuchar el audio que va a responder el humanoide por medio del agente, entonces para que la interacción con el humanoide sea mas real debe activarse unos servomotores que simulan los movimientos de la cabeza, boca y cuello del robot. Estos servomotores son activados con un programa llamado VSA (Visual Servo Automation) que deberá sincronizar los

movimientos con el habla del personaje, brindando la información requerida por el usuario en una forma más real.

3.10.1. Integración con VSA (Visual Servo Automation)

Es un programa que controla y ajusta directamente los servo-motores para una ejecución automática. El funcionamiento del Software VSA es el siguiente:

En la ventana principal se encuentran tres partes importantes como son:

1. Barra de acontecimiento
2. Línea de visión del tiempo
3. Línea de visión de onda

La barra de acontecimiento el usuario define la posición y tiempo de ejecución de un movimiento en cada uno de los servomotores. En la línea de visión de tiempo se observa el inicio y fin de cada movimiento del servo, y en la línea de visión de onda se observa gráficamente las ondas producidas por el audio facilitando la sincronización con el archivo de audio.(Ver **Figura 14**)

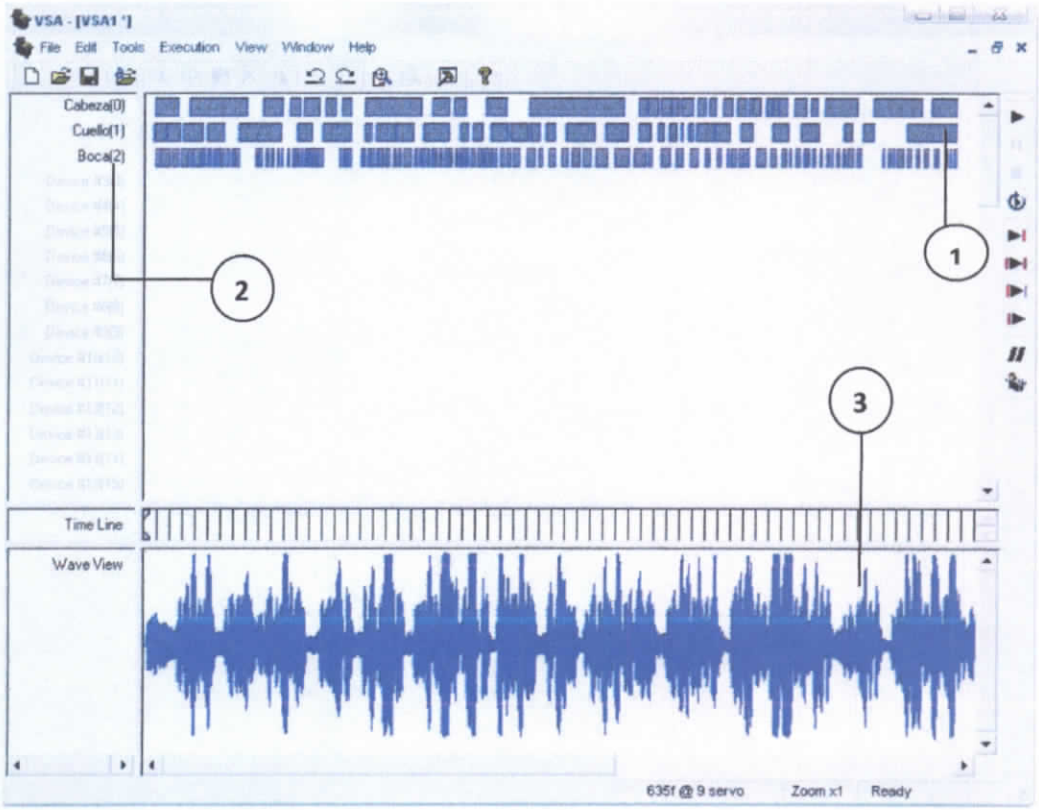


Figura 14: Entorno del VSA

En el proyecto un archivo VSA no va a contener un archivo de audio, ya que el audio va a proporcionar el Agente de Windows, lo único que va a contener el archivo VSA son los movimientos de boca, cuello y cabeza, en un determinado tiempo, por lo tanto para cada frase dependiendo el numero de palabras va a ejecutar un archivo VSA. Por ejemplo si la frase es de tipo "1" va a realizar movimientos de boca por un lapso de 5 segundos, y así sucesivamente el tiempo determinado para cada tipo de frase dependiendo el número de palabras.

3.10.2. Integración con Dragon Naturally Speaking

Los sistemas de reconocimiento de voz en este caso Dragon Naturally Speaking, se enfoca principalmente en las palabras y los sonidos que distinguen una palabra de la otra en un idioma. Estas son los fonemas. Por ejemplo, "tapa", "capa", "mapa", "napa", son palabras diferentes puesto que su sonido inicial se reconoce como fonemas diferentes en nuestro Idioma.

Existen varias maneras para analizar y describir el habla de las personas y definir sus patrones de voz. Los enfoques más comúnmente usados son:

1. **Articulación:** Análisis de cómo el humano produce los sonidos del habla.

2. **Acústica**: Análisis de la señal de voz como una secuencia de sonidos.

3. **Percepción Auditiva:** Análisis de cómo el humano procesa el habla.

Los tres enfoques proveen ideas y herramientas para obtener mejores y más eficientes resultados en el reconocimiento de voz.

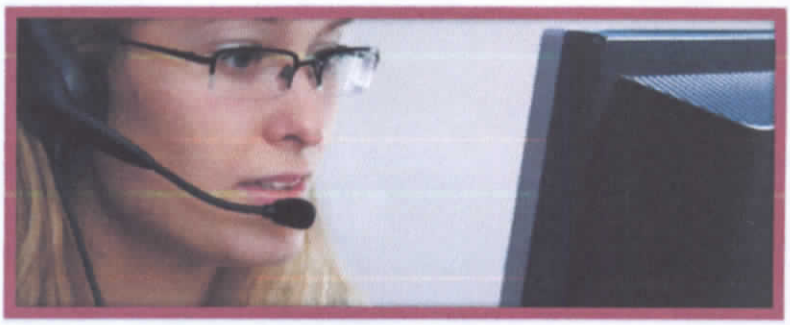
La ventaja de trabajar con patrones de voz es que no necesita analizar características espectrales de la voz a nivel fonético. Lo que nos evita comparar los sonidos formantes y distintivos de cada persona como el tono de voz. Esta comparación resultaría bien si se estuviera comparando un número reducido de palabras. Por ejemplo para comparar la palabra

"Ingeniería en Sistemas" deberíamos pronunciar tal cual se lo hizo en la grabación, con la misma velocidad, y con el mismo tono de voz de principio a fin.

3.10.2.1. Creación de un Usuario

A continuación se detalla la utilización de Dragon Naturally Speaking en la creación de un usuario:

- Colocar el micrófono correctamente, ya que es uno de los factores de precisión en el dictado.



- Calibrar el volumen, ya que permite adaptar el volumen de voz de los usuarios que van a utilizar el programa. Este entrenamiento se lo hace leyendo un párrafo inicial, para esto debe estar colocado correctamente el micrófono, como se recomendó anteriormente.

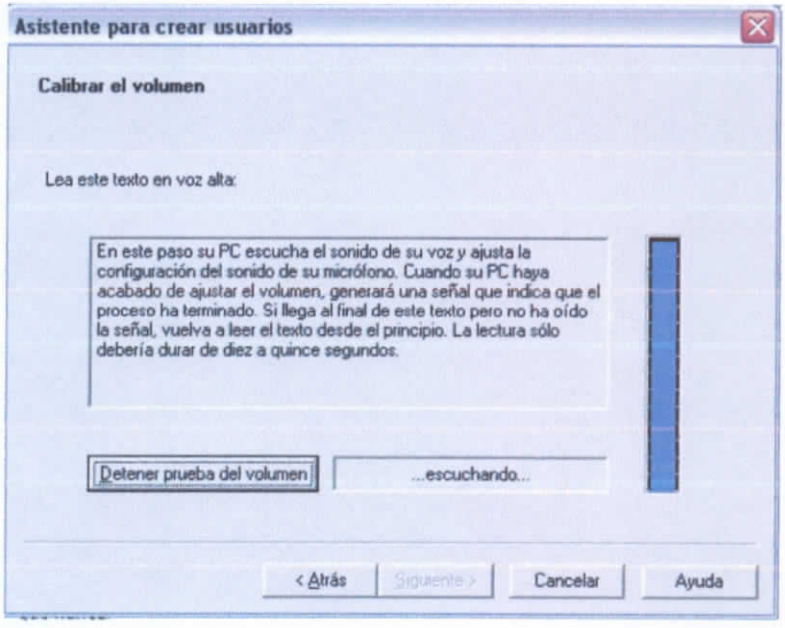


Figura 15. Asistente de Dragon

- Comprobación de la Calidad del sonido. En este paso el software comprueba la calidad del sonido ya sea de la maquina como del micrófono, esta prueba también es importante para conocer si el PC del usuario ayuda a reconocer con claridad las palabras, evitando la mala precisión del reconocimiento.



Si existiera algún inconveniente en este paso, el software emitirá un error, indicando al usuario cual es la falla para corregirla, y sino se resuelve este paso puede pasar al siguiente, pero el sistema advierte que el reconocimiento va a ser impreciso.

- Luego de este paso Dragon prepara los archivos de audio para el usuario. Esto quiere decir que cada sesión de usuario va a tener diferente pronunciación y tono de voz, permitiendo que se vaya guardando esta configuración en el software.

- Una vez completados los pasos de instalación, el software nos da a escoger tres pasos, que dependiendo del usuario escogerá uno de ellos para completar e inicializar Dragon Naturally Speaking.

Los pasos son los siguientes:

- Iniciar el tutorial del programa
- Las Novedades de la versión instalada, en este caso es el Dragon Naturally Speaking en su versión 9 Preferred.
- Empezar con el dictado al software.

- Inmediatamente después de terminar con la creación del usuario, se va a habilitar la barra del software en la parte superior de la pantalla, accediendo de esta manera a la configuración del mismo.

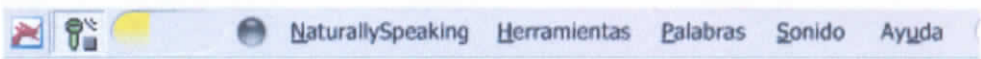


Figura 16. Barra Principal de Dragon

Si la barra se presenta como en el gráfico anterior, ya se puede comenzar a dictar en cualquier programa de Windows que sea editor de texto.

(Nota: Si se desea que el programa únicamente reconozca la voz de un usuario se sigue los pasos anteriores de entrenamiento. La ventaja de la nueva versión de Dragon Naturally Speaking en su versión 9, evita el entrenamiento previo del usuario permitiendo el reconocimiento de voz de varias personas).

3.10.2.2. Centro de Precisión

Ayuda y garantiza al usuario que el reconocimiento de voz sea preciso, para esto Dragon presenta una serie de herramientas de ayudan a mejorar el reconocimiento. Como son:

- **Configuración del sonido**, que indica si se está utilizando de forma correcta el micrófono, además de controlar el volumen del micrófono para que no haya distorsión.

- **Optimizador acústico y del Lenguaje**, que adapta los sonidos emitidos por el usuario y además agrega secuencias de palabras de uso común del modelo del lenguaje.

- **Entrenamiento adicional,** permite entrenar la sesión de usuario y adaptar el tono de voz para cada usuario, por lo tanto si es que se quiere trabajar con varios usuarios en una sesión no es recomendable seguir con el entrenamiento.

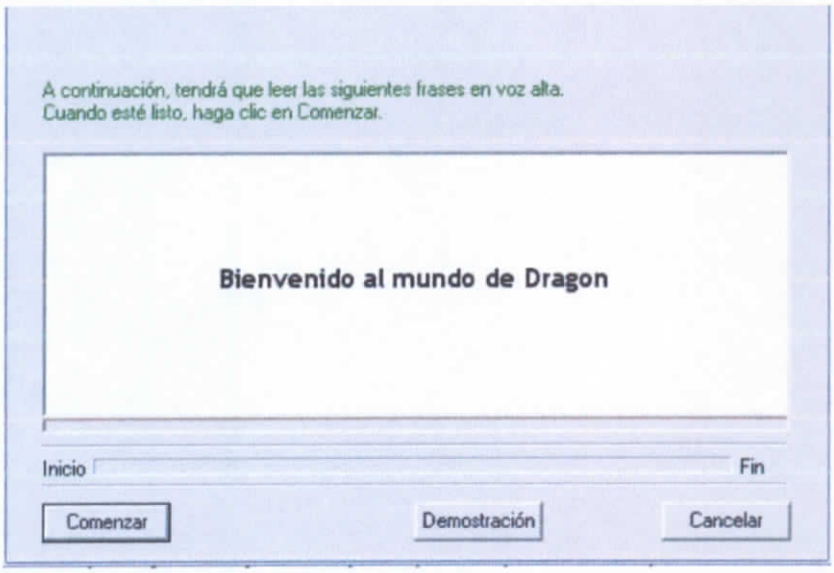


Figura 17. Ventana de Inicio del Entrenamiento de Usuario

- **Visualizador de comandos**, aquí se establecen los comandos que reconoce el programa para inicializar un programa, como por ejemplo si se pronuncia "Iniciar Dragon Pad", automáticamente inicia el editor de texto del programa. Existen otros comandos que hacen tareas específicas accediendo al menú del programa, etc.

- **Añadir nuevas palabras al vocabulario,** para añadir palabras nuevas al usuario, en la barra de Dragon se accede al menú Palabras, dar clic y escoger Añadir, automáticamente sale una ventana la cual permite

ingresar la palabra ya sea escribiendo o dictando por medio del micrófono, que es el medio esencial para el reconocimiento de voz.

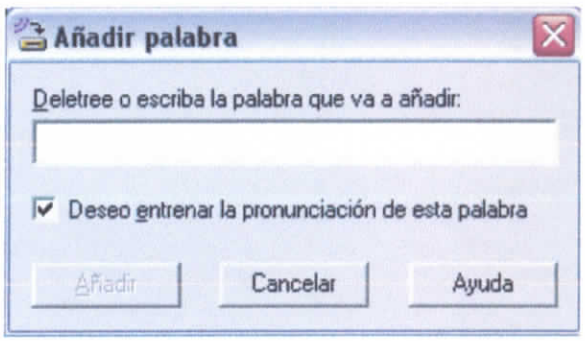


Figura 18. Ventana para añadir palabras nuevas

Una opción importante dentro de este ingreso es la opción de entrenar la pronunciación de la palabra.

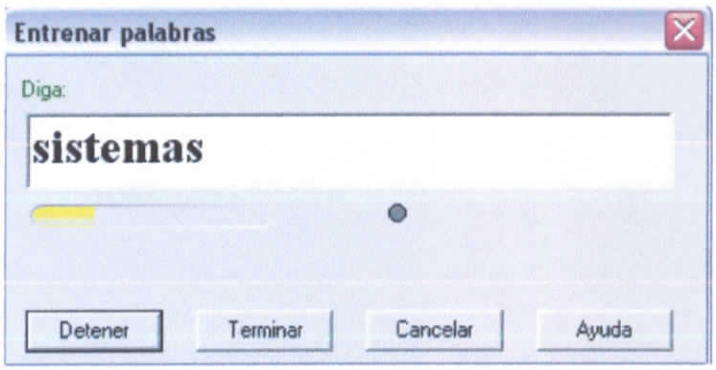


Figura 19. Ventana de entrenamiento de palabras

Este proceso permite reconocer la palabra específicamente para un usuario, el que ha entrenado esta palabra. Como se dijo anteriormente, no es necesario el entrenamiento de las palabras ya que no queremos que se

adapte solo a una persona, esta es la ventaja de esta versión del software permitiendo adaptar las palabras a muchos usuarios.

3.11. Aparato Fonador y medios de Interacción

Los seres humanos poseen cuerdas vocales que permiten comunicarnos emitiendo sonidos las cuales conforman el aparato fonador, el mismo que es identificativo en cada persona. Entonces depende de la persona, los sonidos que produce, y como puede el programa reconocer la palabra para cada uno de los sonidos que se emitan.

Para mejorar la precisión del reconocimiento de voz se debe contar con un micrófono de buena calidad para reducir al máximo los ruidos y eliminar la distorsión del habla.

Otro medio de interacción importante para establecer comunicación con el usuario son los parlantes, que necesariamente deben estar configurados de forma tal que el audio de salida sea normal para establecer una conversación normal con el humanoide.

3.12. Diálogo

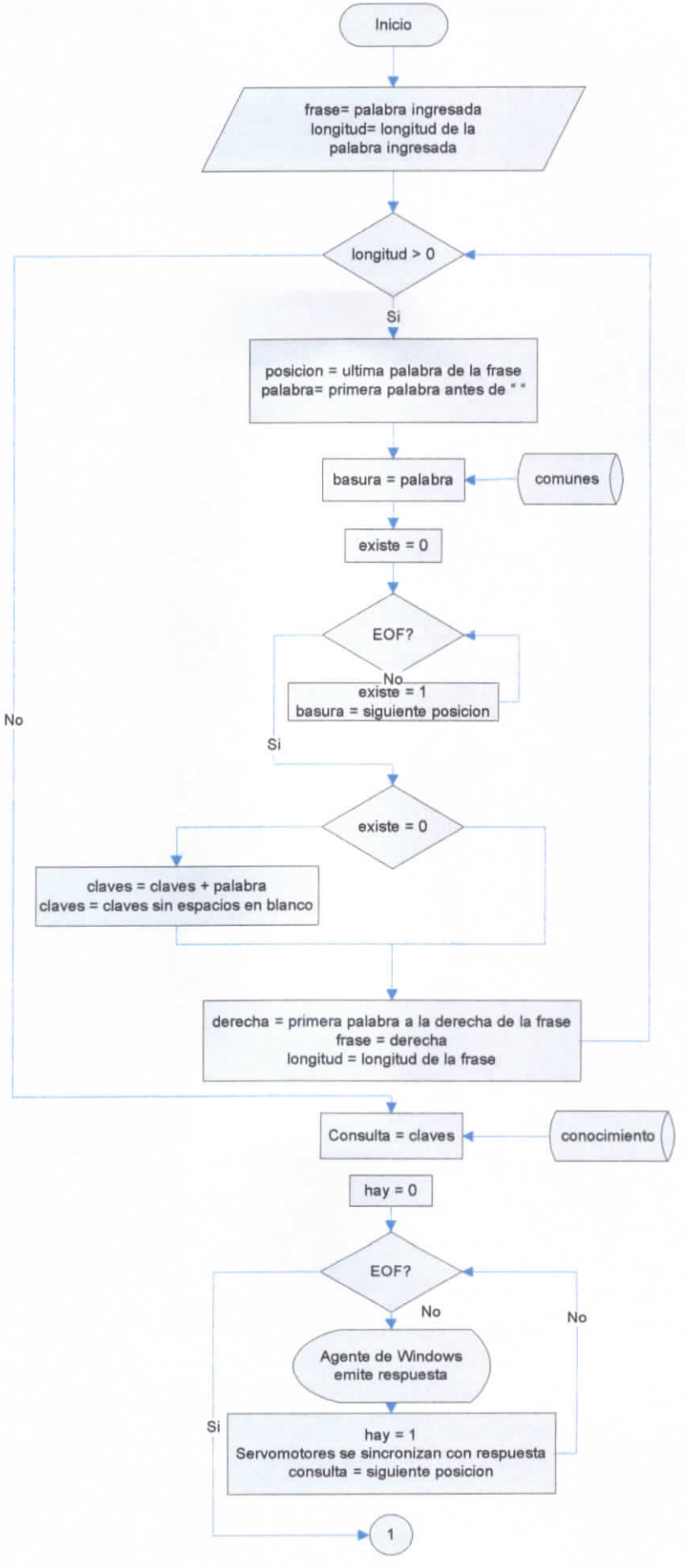
El dialogo que se va a establecer con el humanoide es de forma natural, ya que se toma en cuenta el modismo, la fonética y los formantes de las

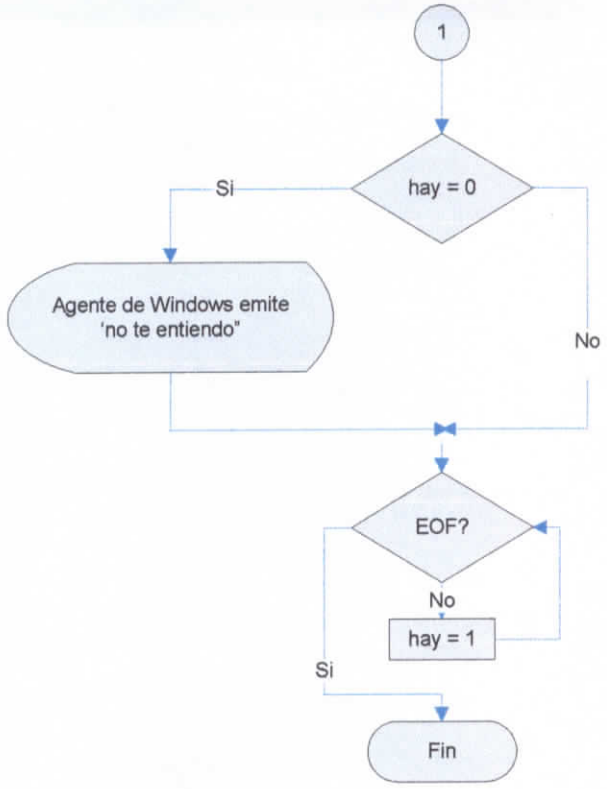
personas, por lo tanto se trata de entablar una comunicación fluida que permita satisfacer las necesidades de la carrera como un buen informante de la misma.

No hay que olvidar que el reconocimiento de voz aún está en sus etapas iniciales ya que como se citó anteriormente se torna dificultosa la tarea de reconocer el sonido y frecuencia del habla de cada persona.

3.13. Control

Un vez establecidos los parámetros de reconocimiento de voz, es necesario controlar lo que se está ingresando o comunicando al humanoide de forma oral. Como necesitamos que el humanoide nos responda las palabras claves ingresadas existe un analizador léxico que permite extraer esas palabras fundamentales para establecer la comunicación. Como hay un sin número de respuestas por ejemplo para "Ingeniería de Sistemas", el humanoide obviamente nos va a dar la información general de la carrera, pero si se le pregunta "Cuantos semestres tiene la Ingeniería de Sistemas", la información que debería brindar al usuario seria más puntual. Por lo tanto es necesario utilizar este analizar léxico para que nos permita rescatar estas palabras puntuales que diferenciarían de diferentes preguntas para dar mejores respuestas. En el caso de que el humanoide no entienda lo que dice o pronuncia el usuario, va a mandar una advertencia de "No entiendo repíteme por favor".


Inicio
frase= palabra ingresada
longitud= longitud de la
palabra ingresada
longitud > 0
Si
posicion = ultima palabra de la frase
palabra= primera palabra antes de " "
basura = palabra
comunes
existe = 0
EOF?
No
existe = 1
basura = siguiente posicion
No
Si
existe = 0
claves = claves + palabra
claves = claves sin espacios en blanco
derecha = primera palabra a la derecha de la frase
frase = derecha
longitud = longitud de la frase
Consulta = claves
conocimiento
hay = 0
EOF?
No
No
Agente de Windows
emite respuesta
Si
hay = 1
Servomotores se sincronizan con respuesta
consulta = siguiente posicion
1



El reconocer la voz de las personas por medio de hardware o software es una tarea muy compleja, además convertir en texto el habla estándar es una tarea extraordinariamente difícil, dada la amplia variación en los patrones del habla humana. Por esta razón, casi todo el software de este tipo (y también algunas de las aplicaciones de comandos de voz básicos) deben ser "entrenadas" para entender la voz de un usuario en particular. Este entrenamiento se lo hace leyendo a la computadora muestras (artículos) de texto preparadas que se incluyen con el software. En cuanto al hardware de reconocimiento o tarjeta de voz, hay que grabar patrones y rango de voces para en una base de patrones y se tornaría una tarea muy complicada pero que a la larga puede reconocer la voz de muchos usuarios tarea que es

difícil para un software de reconocimiento de voz por que únicamente se acostumbra a la voz de un usuario en una sesión.

Entonces si se trabaja con un software de reconocimiento de voz, es necesario entrenar esa sesión de usuario, con diferentes voces para que la tarea de reconocimiento sea más general y sea útil para cualquier tipo de usuario.

CAPITULO IV.

VERIFICACION Y VALIDACION DE RESULTADOS

4.1. Análisis General

El Reconocimiento de voz en los últimos años ha venido evolucionando a cargo de empresas que se han dedicado exclusivamente a perfeccionarse en este tema, dando facilidades a las personas en transcribir en forma oral textos y almacenarlos en la computadora. Tomando en cuenta estas premisas se ha visto necesario implementar esta tecnología en un proyecto de tesis que consta de una cabeza robot en forma de humanoide que brinda información general de la Universidad pero que no es interactiva con los usuarios y que en esta oportunidad se va a tratar de vincular a las personas (usuarios) con el robot humanoide.

El programa necesario y eficaz para el reconocimiento de voz es el Dragon Naturally Speaking que recepta toda la información que es ingresada por el usuario y la convierte en texto para poder ser comparada mediante un programa que busca en una base de datos que contiene almacenada información referente a la carrera de Sistemas de la Universidad Católica. Si esta información es encontrada rápidamente el programa emite la respuesta respectiva.

La información recolectada para almacenar la base de datos, ha sido refinada y seleccionada para satisfacer las necesidades puntuales de los usuarios.

4.2. Pruebas del Proyecto de Tesis

Las pruebas del proyecto permiten probar que el programa desarrollado cumpla con las funciones específicas para las cuales ha sido creado. Por lo tanto la prueba de caja negra que se ha tomado en cuenta en este proyecto ayuda a conocer de mejor manera las opciones principales con las que funciona el humanoide.

4.2.1. Prueba de Caja Negra

- **Opciones de Entrada**

Micrófono:

Lógica:

Opción Válida: volumen alto, volumen bajo

Opción Inválida: Apagado

Rango:

Opción Válida: alto, bajo

Opción Inválida: fuerte, débil

Como se toma en cuenta las opciones iniciales del proyecto a continuación se detallan las pruebas de los botones de tablero de control y de los sensores, respectivamente.

Botones:

Lógica:

Opción Válida: 1, 2, 3, 4

Opción Inválida: Ausente

Rango:

Opción Válida: 1, 2, 3, 4

Opción Inválida: -∞...0

Opción Inválida: 5... ∞

Sensores:

Lógica:

Opción Válida: derecho, izquierdo

Opción Inválida: apagados

Rango:

Opción Válida: derecho, izquierdo

Opción Inválida: arriba, abajo

Opción Inválida: adelante, atrás

4.2.2. Grafos de Causa – Efecto

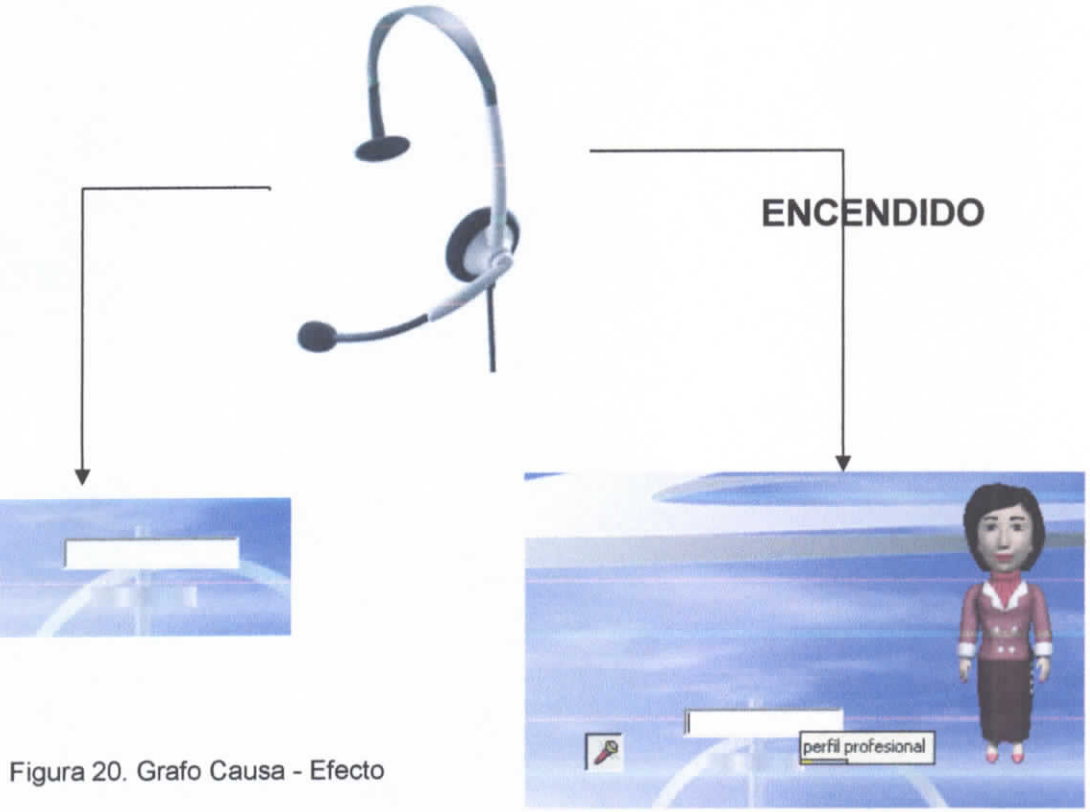


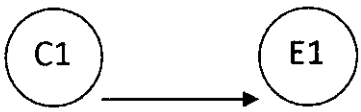
Figura 20. Grafo Causa - Efecto

CAUSAS	
C1	APAGADO
C2	ENCENDIDO

Tabla 5. Tabla de Causas

EFECTOS	
E1	IMPOSIBLE INTERACTUAR
E2	INICIO DE CONVERSACION

Tabla 6. Tabla de Efectos





La pantalla principal del proyecto en el cual se encuentra el agente y el cuadro de texto, no va a ser visible para el usuario, siendo el micrófono el único medio de comunicación con el robot (humanoide).

CONCLUSIONES

- Este proyecto de tesis mediante el humanoide inteligente reconoce la voz humana y brinda información a los usuarios acerca de la Escuela de Ing. de Sistemas de la PUCESA de una manera interactiva y entretenida.

- La información que los usuarios desean obtener del humanoide inteligente se presenta de forma rápida y oportuna.

- El actual proyecto permite a las personas vincularse directamente con la tecnología y la información actual con la que se cuenta, ya que son los beneficiaros directos de este importante avance.

- El reconocimiento de voz se proyecta a futuro como una mejora significativa que causa gran interés en las personas especialmente en los estudiantes que sienten apego a la tecnología como es el reconocimiento de voz y la robótica.

- La sincronización de la conversación usuario – máquina, es una tarea compleja ya que debe ser exacta para que simule el habla natural de una persona.

RECOMENDACIONES

- La conversación del usuario con el humanoide debe ser fuerte y clara para que sean ingresadas las palabras correctas y así obtener la información correcta.

- Investigar acerca de las nuevas tecnologías del reconocimiento de voz y así explotar las ventajas de la misma.

- Una mejora para el humanoide seria trabajar a futuro con el reconocimiento de imágenes ya que el humanoide cuenta con una cámara incorporada que ayudaría a la comunicación con el usuario y perfeccionaría la misma.

- Seria un gran aporte si en futuras mejoras para este proyecto permita al humanoide aprender por sí solo y almacenar cualquier tipo de información referente la Universidad.

- Para trabajar con el reconocimiento de voz es necesario utilizar un micrófono de alta calidad para precisar el ingreso de las palabras.

GLOSARIO DE TERMINOS

A.

ADC.-

Analog to Digital Converter. Es un Conversor analógico-digital. Se encarga de transformar la señal de sonido analógica en su equivalente digital.

API.-

Application Programming Interface (Interfaz de Programación de Aplicaciones).

C.

CPU.-

La unidad central de procesamiento, CPU (por sus siglas del inglés *Central Processing Unit*), o, simplemente, el procesador, es el componente en una computadora digital que interpreta las instrucciones y procesa los datos contenidos en los programas de computadora.

D.

DAC.-

Digital to Analog Converter. Es un Conversor digital-analógico. Su misión es reconstruir una señal analógica a partir de su versión digital.

DBMS.-

Database Management System. Sistema de Gestión de Base de Datos o SGBD.

DI.-

Es el director de la información o vicepresidente de la información que aboga en la institución por los sistemas de bases de datos.

DMA.-

DMA significa *Direct Memory Access*, una capacidad en modernas computadoras que permite a los dispositivos periféricos para enviar datos a la placa base la memoria sin la intervención de la CPU.

DSP.-

Digital Signal Processor. Es un Procesador de señal digital. Es un pequeño microprocesador que efectúa cálculos y tratamientos sobre la señal de sonido, liberando así a la CPU de ese trabajo.

F.

FM.

Frecuencia modulada, un modo de transmisión de datos a través de radio. tipo de tecnología utilizado en tarjetas de sonido de gama media, consistente en reproducir el sonido mediante un sintetizador musical FM, obteniendo un resultado menos real que el ofrecido por las tarjetas para el reconocimiento de voz.

I.

ISA.-

El *Industry Standard Architecture* (en inglés, Arquitectura Estándar Industrial), casi siempre abreviado ISA, es una arquitectura de bus creada por IBM en 1980 en Boca Raton, Florida para ser empleado en los IBM PCs.

M.

MIDI.-

Son las siglas de *Musical Instrument Digital Interface* (Interfaz Digital de Instrumentos Musicales). Se trata de un protocolo industrial estándar que permite a las computadoras, sintetizadores, secuenciadores, controladores y otros dispositivos musicales electrónicos comunicarse y compartir información para la generación de sonidos.

N.

NLC.

Natural Language Command, Comandos de Lenguaje Natural.

O.

OCX.-

"*OLE Control Extension*". Son controles y funciones para ser utilizados en programas para Windows.

OLE.-

Object Linking and Embedding. Tecnología de Microsoft que permite vincular elementos a diferentes aplicaciones.

P.

PC.-

Personal Computer (Computadora Personal).

PCI.-

Un *Peripheral Component Interconnect* (PCI, "Interconexión de Componentes Periféricos") consiste en un bus de ordenador estándar para conectar dispositivos periféricos directamente a su placa base.

PLN.-

Procesamiento del Lenguaje Natural. Es una disciplina muy amplia que, además de la Recuperación y organización de la Información incluye campos como la traducción automática, reconocimiento del habla, corrección ortográfica de textos, etc.

R.

RCA.-

El conector RCA es un tipo de conector eléctrico común en el mercado automotor. El nombre "RCA" deriva de la *Radio Corporation of America*, que introdujo el diseño en los 1940.

ROM.-

Read-Only Memory, que significa "memoria de sólo lectura": una memoria de semiconductor destinada a ser leída y no destructible, es decir, que no se puede escribir sobre ella y que conserva intacta la información almacenada, incluso en el caso de que se interrumpa la corriente (memoria no volátil). La ROM suele almacenar la configuración del sistema o el programa de arranque de la computadora.

S.

S/PDIF.-

El estándar S/PDIF (*Sony/Philips Digital Interface*, o abreviado *SPDIF*) es un formato digital de transferencia de datos de audio.

SAPI.-

Speech Application Programming Interface. Es una API desarrollada por Microsoft el cual permite usar el reconocimiento de voz para todas las aplicaciones de Windows.

SDK.-

Un *Software Development Kit* (SDK) o kit de desarrollo de software es generalmente un conjunto de herramientas de desarrollo que le permite a un programador crear aplicaciones para un sistema bastante concreto, por ejemplo ciertos paquetes de software, frameworks, plataformas de hardware, ordenadores, videoconsolas, sistemas operativos, etcétera.

SQL.-

Structured Query Language (Lenguaje de consulta estructurado). Es un lenguaje declarativo de acceso a bases de datos relacionales que permite especificar diversos tipos de operaciones sobre las mismas.

T.

TIC.-

Tecnologías de la información y la comunicación.

TOKENS.-

Tokens son los elementos en que el preprocesado desmenuza el código fuente. En un lenguaje de programación. Los tokens son el equivalente a las palabras y signos de puntuación en el lenguaje natural escrito. Los tokens están separados por elementos de separación que reciben el nombre genérico de separadores.

U.

USB.-

El *Universal Serial Bus* (bus universal en serie) o Conductor Universal en Serie es un puerto que sirve para conectar periféricos a una computadora.

V.

VSA.-

Visual Servo Automation (Automatización visual de Servos).

BIBLIOGRAFIA

LIBROS

- Aníbal Ollero Baturone "ROBOTICA "Manipuladores y robots móviles"; Marcombo ,Barcelona – España, 2001.

- James L. Johnson 'Bases de Datos, Modelos, lenguajes, diseño". USA; OXFORD, 2000.

- J. González B. "Electrónica General" Tomo V. Madrid - España; 1983

- HESS ARAYA, "Inteligencia artificial y derecho", en Revista Electrónica de Derecho e Informática numero 39, octubre del 2001.

- Marvin Minsky, "Robótica: La ultima frontera de la alta tecnología, Planeta, Barcelona , 1985.

- Torres, F; Pomares, J; Gil, P; Puente, S.T Aracel R. "Robótica Automática: Robots y Sistemas Sensoriales" Prentice Hall, 2002. ISBN: 84-205-3574-5.

- Nils J. Nilsson "Inteligencia Artificial: Una nueva Síntesis", Mc.Graw Hill, 2001.

- Daniel Crevier "Inteligencia Artificial: Ordenadores y Programación" (Acento Editorial), 1996.

Libros Electrónicos

- Departament of Electronics and Tele-communication Engineering Jadavpur University, Calcuta, India, Konar Amit "Artificial Intelligence

and Soft Computing: Behavioral and Cognitive Modeling of the Human Brain", 1999. Cuanto libro

ISBN: 0-8493-1385-6

http://www.cuantolibro.com/libro/22145/Artificial-Intelligence-And-Soft-Computing.html

INTERNET

Robótica

http://es.wikipedia.org/wiki/Rob%C3%B3tica

Automatización

http://www.sc.ehu.es/sbweb/webcentro/automatica/WebCQMH1/PAGINA%20PRINCIPAL/Automatizacion/Automatizacion.htm

Tareas Programadas

http://lmsextremadura.educarex.es/file.php/33/Robots/robots.htm

Base de datos

http://www.monografias.com/trabajos7/bada/bada.shtml

Electrónica

http://es.wikipedia.org/wiki/Electr%C3%B3nica

Reconocimiento de voz - Hardware

http://ict.udlap.mx/people/ingrid/Clases/IS412/index.html

http://www.mailxmail.com/curso/informatica/hardwareaudio/capitulo20.htm

http://www.spanish-translator-services.com/espanol/preguntas-frecuentes/Reconocimiento-de-voz-hardware.htm

Reconocimiento de voz - Software

http://www.itq.edu.mx/vidatec/espacio/aisc/ARTICULOS/ASCC/articuloASCC.htm

Configuración de voz

http://www.mailxmail.com/curso/informatica/hardwareaudio/capitulo22.htm

Tarjeta de sonido

http://es.wikipedia.org/wiki/Tarjeta_de_sonido

Modelado del lenguaje

http://64.233.169.104/search?q=cache:8bbaxB6DmyoJ:eprints.rclis.org/archive/00010024/01/PROCESAMIENTO_DEL_LENGUAJE_NATURAL_EN_LA_RECUPERACI_N_DE_INFORMACI_N.pdf+modelo+del+lenguaje+ia&hl=es&ct=clnk&cd=3&gl=ec

Tarjeta de sonido

http://www.pchardware.org/sonido.php

SOUND BLASTER:

http://es.wikipedia.org/wiki/Sound_Blaster

Motor de Inferencia

http://www.hess-cr.com/secciones/dere-info/asistadmi.shtml

Sistemas Inteligentes

http://www.intelligent-systems.com.ar/intsyst/intsystSp.htm

Reglas IA

http://www.biografica.info/redei/introduccion-a-la-inteligencia-artificial-

22.php?split=22

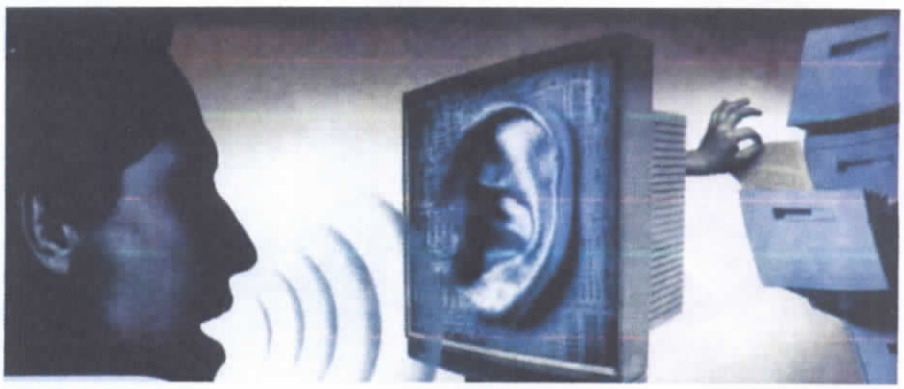
Algoritmo de búsqueda

http://es.wikipedia.org/wiki/Algo

ritmo_de_b%C3%BAsqueda

ANEXO

Manual de Usuario



Robot Inteligente con Reconocimiento de Voz

INTRODUCCION

Aprovechando el mecanismo robótico con el que contaba la Universidad Católica, se ha visto necesario brindar información más específica a los usuarios, permitiéndoles interactuar con el humanoide y así satisfacer las necesidades informativas relevantes de la carrera de Ingeniería de Sistemas.

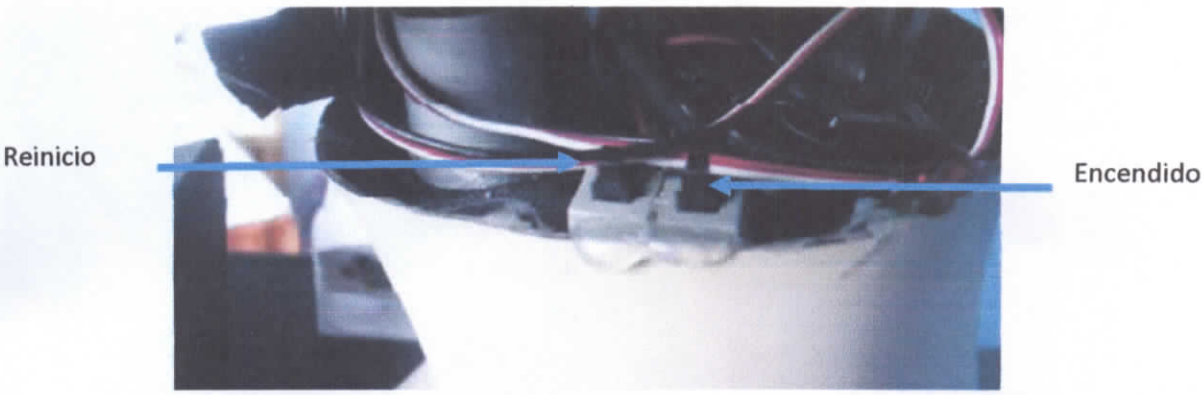
ENCENDIDO DEL ROBOT

Para utilizar el robot (Humanoide) es necesario contar con un cable de poder para conectarlo a la energía eléctrica y así poder encenderlo.



Cable de Poder

El encendido del robot se lo encuentra en la parte del cuello, en donde encontramos dos botones, el uno que sirve para el encendido y el otro para reiniciar el robot.



Botones de encendido y reinicio

Una vez encendido el robot (humaoide) se inicia la aplicación, Dragon Naturally Speaking y el agente de Windows el cual será el principal objeto de interacción con el usuario.



Ventana Principal del Programa

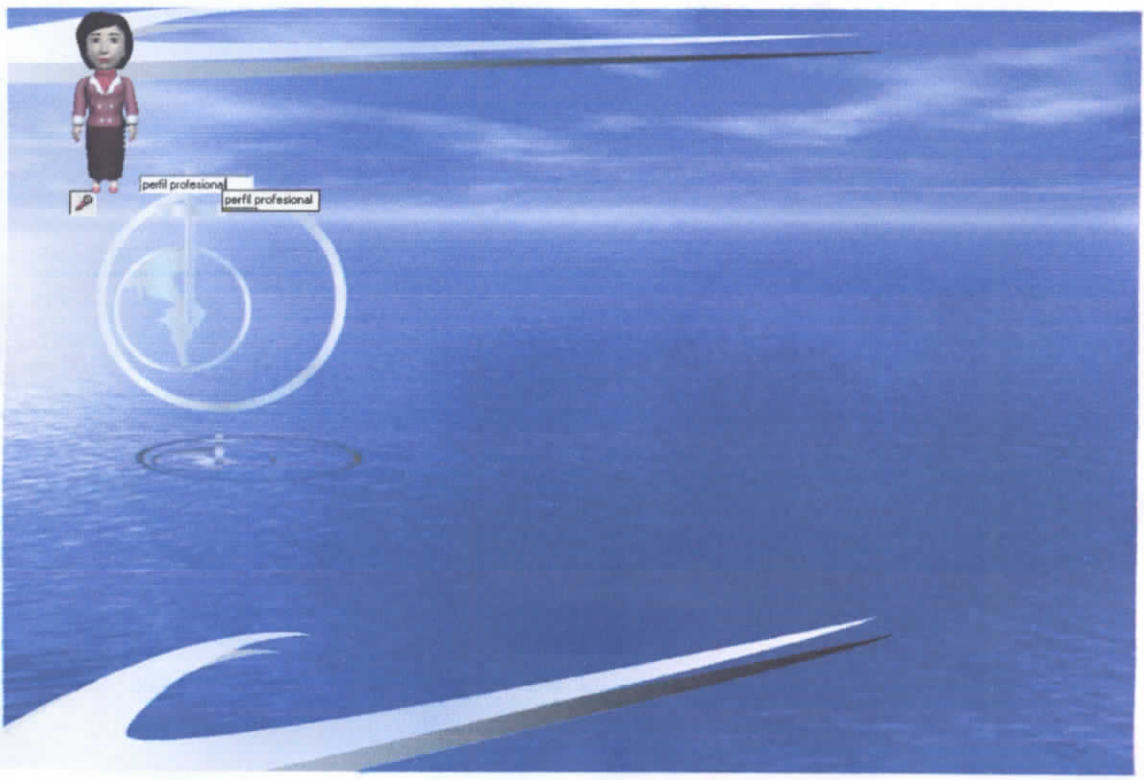
Debemos cerciorarnos de conectar bien y encender el micrófono para comenzar a conversar con el humanoide.

INTERACCION CON EL HUMANOIDE

Una vez hecho lo anterior esta listo el robot para la interacción con el usuario. Para comenzar a conversar es necesario hablar claro y con un tono de voz alto para que el reconocimiento de voz sea más exacto.

Toda la información almacenada en la base del conocimiento del robot, como ya se citó anteriormente, esta centrada en la carrera de Ingeniería de Sistemas por lo tanto el tema de conversación debe estar relacionado a temas que conciernen a esta carrera por ejemplo:

Si se necesita saber acerca del perfil profesional que genera la Universidad con sus estudiantes, solo pronunciamos el comando que contenga en alguna parte de la frase "*Perfil Profesional*" y nos responde según esté almacenada en la base del conocimiento la información requerida.



Comando ingresado por el micrófono

Es así que podemos entablar una conversación con el humanoide, y si la palabra ingresada no es encontrada, es decir si no entiende el robot, sencillamente emitirá una respuesta de: "No entiendo la pregunta, repíteme por favor".

COMANDOS PREDEFINIDOS

Las opciones del tablero de control también se encuentran como comandos para que el humanoide responda información general de la Universidad Católica.

Por lo tanto el usuario debe hablar las siguientes frases para obtener respuesta por parte del humanoide: