

# Interacción humano-máquina por voz para la operación de plataformas robóticas móviles

## **Héctor Gutiérrez Hernández**

Universidad del Istmo, Bo. Sta. Cruz Tagolaba, Tehuantepec, Oaxaca, México, Tel. 9711401747  
*gutierrez.h3ctor@gmail.com*

## **J. Jesús Arellano Pimentel**

Universidad del Istmo, Bo. Sta. Cruz Tagolaba, Tehuantepec, Oaxaca, México, Tel. 9715224050  
*jjap@sandunga.unistmo.edu.mx*

## **Daniel Pacheco Bautista**

Universidad del Istmo, Bo. Sta. Cruz Tagolaba, Tehuantepec, Oaxaca, México, Tel. 9715224050  
*dpachecob@bianni.unistmo.edu.mx*

## **Resumen**

En el presente trabajo se describe el desarrollo de una interfaz de comunicación capaz de reconocer una serie de comandos teniendo como entrada la señal acústica emitida por un usuario con el objetivo de interactuar con plataformas robóticas móviles. Para realizar la comunicación entre la interfaz de usuario y la plataforma móvil se realizó un diseño modular con una arquitectura cliente-servidor. Esta comunicación se implementó mediante sockets orientados a conexión y el software *Player/Stage* que permite controlar los dispositivos de un robot y obtener información de sus sensores. En el sistema de reconocimiento del habla se utilizaron los modelos ocultos de Markov, a través de la plataforma HTK, empleando cinco estados con 3 mixturas Gaussianas por estado; además, se diseñó un corpus de entrenamiento para el español Mexicano utilizando el alfabeto *Mexbet*. Durante los experimentos realizados se obtuvieron porcentajes de reconocimiento de 98.26% en palabras y 94.39% en frases reconocidas de manera completa; la validación de los resultados fue hecha mediante el método “*Word Accuracy*”.

**Palabras Claves:** Reconocimiento del habla, Robots móviles, Modelos ocultos de Markov.

## 1. Introducción

Durante las últimas décadas se ha estudiado la posibilidad de desarrollar interfaces de interacción humano-máquina controlados por la voz para conseguir que las máquinas interactúen con los humanos de manera natural, sin la necesidad de utilizar las interfaces tradicionales basadas en teclado, joystick o dispositivos similares [1]. Por tal motivo han surgido los Sistemas de Reconocimiento Automático del Habla (SRAH), cuyo objetivo es transformar la señal de voz en una representación de la misma, normalmente en formato escrito. Los SRAH suelen clasificarse de manera general en dos tipos: 1) dependientes del usuario, es decir, se construyen tomando en cuenta las características acústicas particulares de un solo usuario y solo él puede usar el sistema, y 2) independientes del usuario, los cuales se construyen a partir de características de voz de muchos usuarios, para posteriormente personalizar el sistema mediante técnicas de adaptación de usuario. Otras formas de clasificarlos es en función del vocabulario soportado, del tipo de locución de entrada (desde palabras aisladas hasta el habla continua) o incluso en función de si el sistema está adaptado o no al locutor [2].

Actualmente se pueden observar grandes avances tecnológicos y aplicaciones que se les han dado a los SRAH, sin embargo, la mayoría de estos sistemas están diseñados para el idioma inglés y muy pocos para el español mexicano. En base a los sistemas de reconocimiento de voz se puede mencionar el trabajo [3], donde los autores diseñan un sistema de diálogo cuya finalidad es entender telefónicamente a los usuarios y ofrecer información sobre cines. Para el reconocedor del habla, usaron la biblioteca HTK y fue entrenado para identificar gramáticas para el español peruano. Otro trabajo se muestra en [4], donde se plantea una solución que permite la identificación de comandos de voz, con un diccionario reducido, para identificar comandos: adelante, atrás, derecha, izquierda y pare. Para el reconocimiento de los comandos utilizaron clasificadores estocásticos GMM y HMM.

En base al reconocimiento de voz con aplicación para la robótica móvil desarrollados en México se puede mencionar el proyecto DIME (Diálogos Inteligentes Multimodales en Español) desarrollado en el Departamento de Ciencias de la Computación del Instituto de Investigación en Matemáticas Aplicadas y Sistemas (DCC-IIMAS) de la UNAM, en el cuál se emplea SRAH de elaboración propia que permite la interacción por medio de lenguaje natural entre el robot y los usuarios humanos por medio de diálogos cognitivos con el sistema [4,5]; el principal objetivo del robot, además de conversar con el usuario, es proporcionar una visita guiada por el DCC-IIMAS. La versión mejorada tiene el objetivo de servir a los seres humanos en sus tareas diarias [6].

Otro proyecto desarrollado en México es el robot de servicio llamado Sabina, el cual es una mejora del proyecto Markovito [7] desarrollado en el INAOE, en el cual se emplea en una plataforma robótica PatrolBot y los módulos se integran utilizando una arquitectura basada en comportamiento en capas que utiliza memoria compartida para la comunicación. El propósito del proyecto es obtener un mejor rendimiento de la percepción e interacción efectiva con las personas, además, tiene planeado que sirva de apoyo en la rehabilitación de pacientes [8].

A diferencia del proyecto DIME o Sabina, el presente trabajo se enfoca en el desarrollo de una interfaz donde sea el usuario quién determine los movimientos de la plataforma robótica móvil a través de comandos de voz, así el robot se moverá en su entorno a voluntad del operador siempre y cuando no exista el peligro de una colisión inminente al ejecutar una orden. Además, mediante un sintetizador de voz el robot interactúa con el usuario para confirmar la orden recibida e informarle si es capaz de ejecutar dicha orden o no, todo esto de una manera natural y amigable como si el diálogo fuese directamente con otra persona. Así, la aplicación final del sistema propuesto está enfocada para que personas con discapacidades motrices puedan operar, mediante comandos de voz, una silla de ruedas robotizada. De esta manera la persona con la discapacidad se sentará en la silla y por medio de órdenes vocales le indica a ésta, el sentido de los giros o movimientos deseados. Sin embargo, dado que aún se está trabajando sobre el sistema de control de la silla de ruedas, las pruebas se han desarrollado con un robot móvil ERA-MOBI como prueba piloto previo a la operación en la silla de ruedas.

## 2. Desarrollo

Para la interfaz de comunicación propuesta se consideró operar un sistema robótico de la manera más parecida al lenguaje natural. Dado que las actividades del sistema robótico son limitadas se restringirá el conjunto de frases reconocibles por la interfaz a un modelo de lenguaje específico mejorando también así el resultado del reconocimiento. Cabe mencionar que el desarrollo del proyecto se ha dividido en tres fases: el sistema de reconocimiento automático del habla, el diseño visual de la interfaz y el proceso de comunicación entre la interfaz y la plataforma robótica móvil.

### 2.1 Esquema general de desarrollo

En la Fig.1 se presenta el esquema general de la interfaz de comunicación por voz realizado. Se observa que la interfaz toma como entrada una secuencia de señales acústicas del usuario, obtenidos desde un micrófono convencional; está es procesada por medio de los Modelos Ocultos de Markov implementados en HTK y genera una frase de operación para el robot, esta frase se interpreta para obtener el código que indica la operación de movimiento del sistema robótico móvil. Posteriormente, se envía el código de la frase a la Interfaz cliente-servidor, lo interpreta y mediante los datos de los sensores monitorea las condiciones de operación del robot móvil para indicarle al usuario si la operación es permitida o no. En el caso de que sea permitida, el Servidor Player recibe las órdenes y conduce la plataforma robótica hasta la posición indicada. Además, la Interfaz Cliente-Servidor genera una respuesta frente a comandos de voz expresados por el usuario y mediante el sintetizador de la interfaz se reproduce la voz artificialmente.

La implementación del sistema se basa en un diseño modular con una arquitectura cliente-servidor, de tal forma que tanto la interfaz como el dispositivo móvil son sistemas totalmente independientes. De esta manera, el sistema bajo Linux puede ser utilizado por cualquier aplicación implementada en cualquier lenguaje compatible con *sockets TCP*;

del mismo modo la interfaz que corre bajo Windows puede ser utilizada para operar cualquier tipo de plataforma robótica móvil bajo los mismos parámetros, así en un futuro el robot ERA-MOBI puede ser reemplazado por una silla de ruedas robotizada.

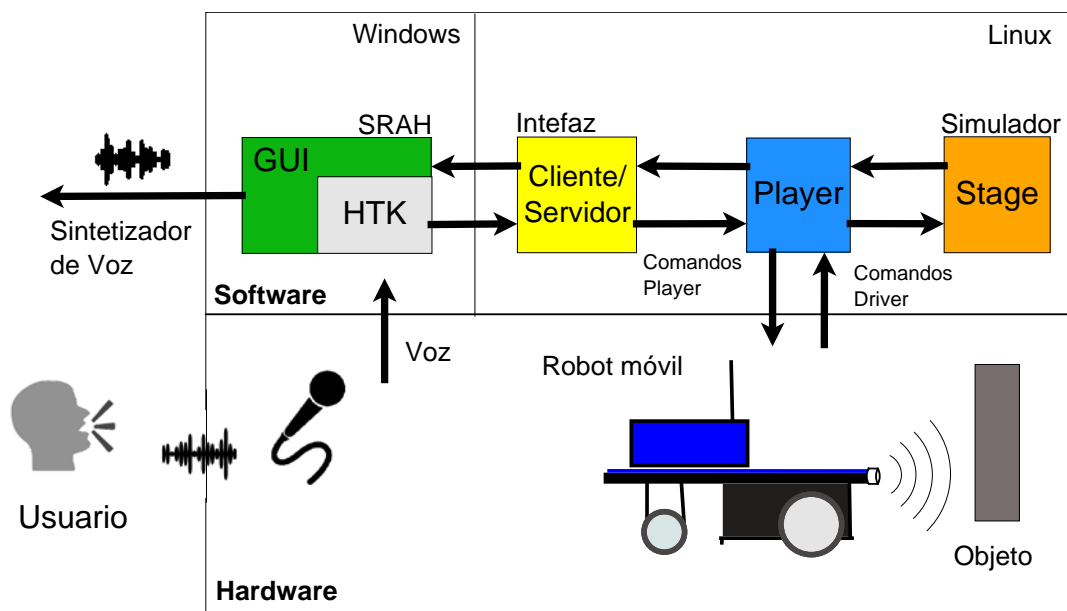


Fig. 1. Esquema general de desarrollo.

A continuación se describe brevemente el desarrollo de la interfaz humano-máquina para operación de la plataforma robótica móvil.

## 2.2 El reconocedor de voz

Los sistemas de reconocimiento del habla consideran que la señal de voz es la realización de algún mensaje codificado como una secuencia de uno o más símbolos. En la Fig. 2 se presentan los componentes base de un SRAH, los cuales son independientes del tipo de sistema e incluyen la adaptación de un nuevo usuario. Este módulo de reconocimiento se realiza mediante un modelo acústico de palabras a nivel fonema, ya que así se modela mejor la coarticulación de las palabras.

De acuerdo con el esquema de desarrollo, como primer paso es necesario construir un corpus de entrenamiento, formado por un corpus textual y un oral. Dicho corpus de entrenamiento, es utilizado para la construcción y entrenamiento supervisado de los modelos acústicos, los cuales son representados mediante los Modelos Ocultos de Markov. Por otro lado, tanto el modelo de lenguaje y el diccionario fonético son contruidos a partir del corpus textual. Una vez construido el modelo de lenguaje, el diccionario fonético y los modelos acústicos, son utilizados por el algoritmo de búsqueda para estimar la palabra más probable ( $\hat{W}$ ) dada una muestra de voz ( $O$ ). Además, la transcripción de referencia (la palabra correcta) se representa mediante la  $W$ .

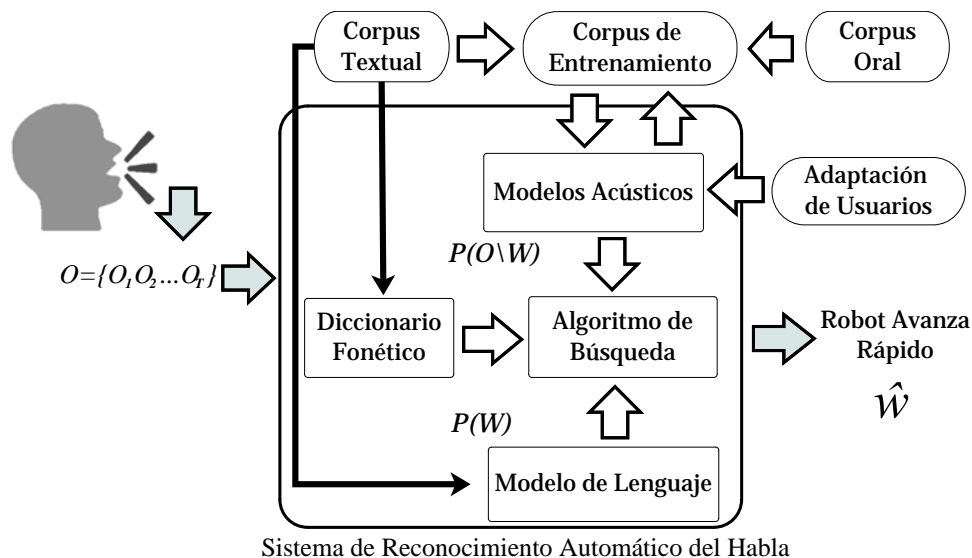


Fig. 2. Esquema general de desarrollo para un SRAH.

### Corpus de entrenamiento

Un corpus de entrenamiento se define como una base de datos o colección de archivos de voz (audio) y transcripciones textuales del mismo en un formato que puede ser usado

para la creación y refinación de modelos acústicos para los sistemas de reconocimiento del habla. Dentro de estos corpus se pueden diferenciar dos tipos:

- Textual. Consiste en una colección de texto representativo de un lenguaje.
- Oral. Representado por una colección de archivos de audio (voz), generalmente obtenidos por la lectura del texto representativo.

Para este trabajo, el corpus de entrenamiento constó de 375 palabras, con las cuales se construyó un corpus textual de entrenamiento con 125 frases. Dichas frases se construyeron de tal manera que incluyeran al menos cinco muestras de voz por cada fonema del español mexicano, lo anterior con base en [9] donde se concluye que con cinco muestras de voz se pueden obtener altos porcentajes de reconocimiento, aún con personas que presentan discapacidades de pronunciación en el habla. Es importante mencionar que los 29 fonemas utilizados en este trabajo (ver Tabla 1) se definieron en base al alfabeto *Mexbet* [10].

Fonemas del Español Mexicano
/a/ /b/ /d/ /e/ /f/ /g/ /i/ /x/ /k/ /l/ /m/ /n/ /nn/ /o/ /p/
/r/ /s/ /t/ /u/ /ts/ /ks/ /Z/ /_R/ /_G/ /_N/ /_D/ /r(/
/sil/ /sp/

**Tabla 1. Fonemas para el español mexicano definido por Mexbet.**

Para obtener el corpus oral, se procedió a realizar la lectura de las frases de entrenamiento. Dicho corpus se genera con las muestras de voz de 10 usuarios diferentes (5 hombres y 5 mujeres), con esto se asegura que las muestras de voz incluyan diferentes variaciones de tono, intensidad de voz y velocidad de pronunciación, además, esto permite contar con un reconocimiento más globalizado. Las frases fueron grabadas con

la ayuda de un micrófono convencional y usando una aplicación implementada en Visual Basic, dichas frases fueron grabadas con una frecuencia de muestreo de 8 kHz monoaural en formato WAV. Cada uno de los archivos de audio es almacenado y codificado para obtener un muestreo de la voz del usuario. La codificación utilizada es la MFCC (*Mel Frequency Cepstral Coefficients*) [11, 12], la cual se basa en la variación conocida de los anchos de bandas de las frecuencias críticas que tiene el oído humano. Para esta codificación se empleó el módulo denominado HCopy de HTK especialmente diseñado para la codificación de voz en MFCC.

### **Modelo acústico y diccionario fonético**

El modelado acústico consiste en el proceso de establecer representaciones estadísticas para las características espectrales de la señal de voz. En este caso, el modelo utilizado para representar las características acústicas de las voces de los usuarios son los Modelos Ocultos de Markov. Dichos modelos se generaron con cinco estados, de los cuales tres son estados emisores y dos estados no emisores (sin función de observación). Además se utilizaron tres componentes gaussianos por estado.

Típicamente, los SRAH se modelan a nivel fonema, pues el fonema es la unidad básica de sonido con significado en un idioma con la que se puede formar una amplia variedad de palabras. Por ejemplo, la palabra IMAGEN se forma por la secuencia de fonemas /i/ /m/ /a/ /g/ /e/ /n/. En este sentido, el diccionario fonético es usado para establecer las secuencias transcripción fonética que forman cada palabra del vocabulario.

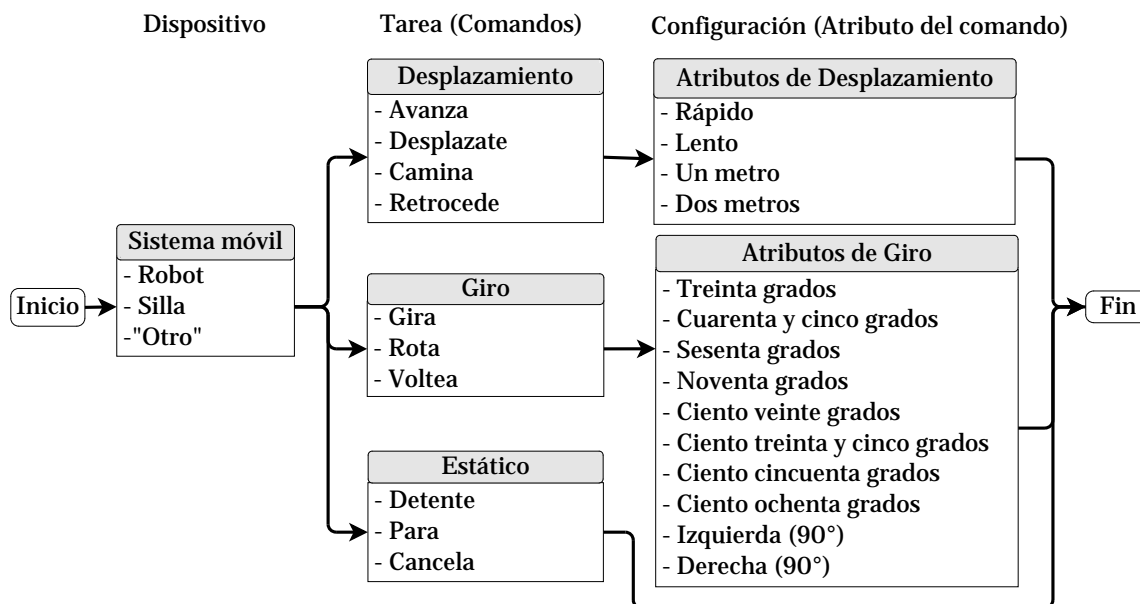
### **Modelo de lenguaje**



Un modelo de lenguaje representa un conjunto de reglas o probabilidades que restringen la secuencia de palabras reconocidas por el SRAH a secuencias más válidas. Para el caso de estudio mostrado en este trabajo, se definieron los comandos de operación del sistema robótico móvil, los cuales deben seguir la siguiente estructura:

*Dispositivo + Tarea + Configuración*

Donde *Dispositivo* define el nombre o identificador del sistema robótico, *Tarea* identifica el tipo de acción que realizará el sistema robótico y *Configuración* proporciona los detalles de la tarea que se desea realizar (ver Fig. 3).



**Fig. 3. Frases para los comandos de operación.**

De este modo, a partir del modelo del lenguaje el reconocedor permite identificar comandos simples de mediana complejidad y mejorar también así el resultado del reconocimiento. Es importante mencionar que tanto el modelo acústico como el modelo del lenguaje fueron implementados utilizando la herramienta HTK.

## **Adaptación**

El desempeño de un SRAH depende de las muestras de voz utilizadas para el entrenamiento, y su desempeño puede ser diferente con distintos usuarios cuyas voces no se usaron para entrenar el sistema. En este trabajo se utilizó la técnica MLLR [13] para implementar la adaptación de nuevos usuarios a un SRAH. MLLR se basa en el supuesto de que un conjunto de transformaciones lineales se puede usar para reducir la diferencia entre los modelos acústicos de un reconocedor de voz y los datos de adaptación. Estas transformaciones son aplicadas sobre la media y varianza de las mixturas de gaussianas de los HMM del sistema base, teniendo el efecto de ajustar dichos parámetros de tal manera que aumente la probabilidad de que los HMM's del sistema generen los datos de adaptación.

### **2.3 Módulos de la Interfaz gráfica de comunicación**

En esta Sección se presenta los principios de diseño e implementación de la interfaz gráfica de comunicación, comentando cada uno de los distintos módulos que lo integran y cómo estos se relacionan entre sí. La interfaz de comunicación se ha dividido en tres módulos principales:

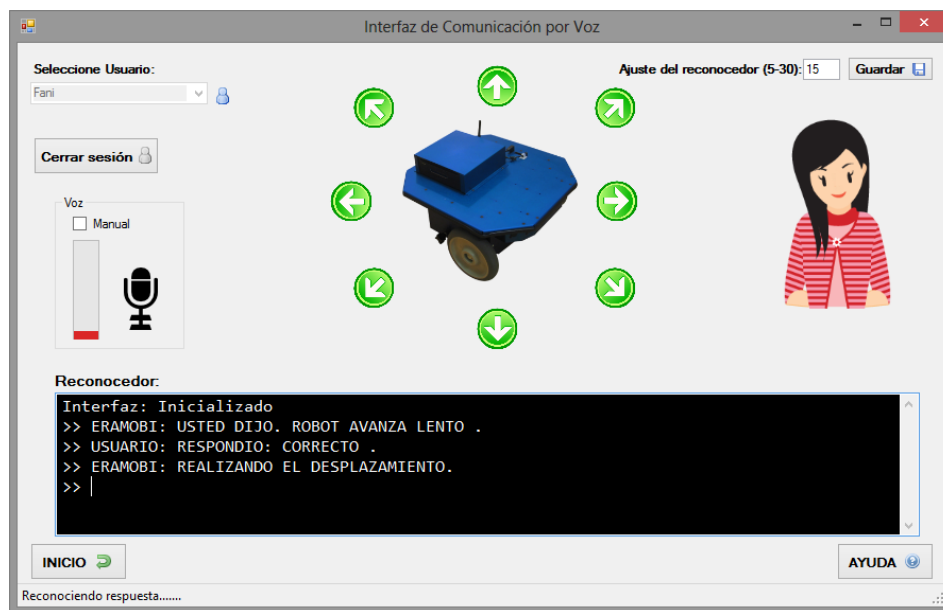
- **Administración de la Interfaz:** permite al usuario personalizar la interfaz, por ejemplo: agregar o eliminar usuarios, configurar los datos necesarios para la conexión con el sistema robótico móvil y seleccionar el tipo de ayudante de la interfaz.
- **Adaptación del Reconocedor de Voz:** proporciona al usuario una rutina de grabación de frases para la adaptación de usuario, para ajustar los parámetros de los Modelos Ocultos de Markov a su forma de voz.
- **Sistema de Reconocimiento de Voz:** permite al usuario operar los movimientos del sistema robótico móvil mediante comandos de voz.

Para el desarrollo de este trabajo se utilizó Visual Basic 2012 (VB, por sus siglas en inglés) como lenguaje de programación y Microsoft Visual Studio Express Edition como entorno de desarrollo integrado (IDE, por sus siglas en inglés) distribuido por Microsoft Corporation. Esta plataforma de desarrollo bajo Windows es de carácter gratuito, orientándose a principiantes y estudiantes.

En este proyecto se consideró a dos tipos de usuarios, el usuario que padece la discapacidad motriz (el que utilizará la interfaz) y otro que no lo padece (configurará la interfaz). Se asume que estos tipos de usuarios conocen los conceptos básicos sobre informática y las tareas de la interfaz. Además, el diseño se basó en los principios de diseño de interfaces de Ben Shneiderman [14].

### **2.3.1 Sistema de Reconocimiento de Voz**

Una vez construido el modelo acústico adaptado para un nuevo usuario, éste ya puede comenzar a usar el reconocedor de voz para operar el sistema robótico móvil. Para ello accede al módulo de la interfaz “Sistema de Reconocimiento de Voz”, el cual se presenta en la Fig. 4.



**Fig. 4. Interfaz del módulo de reconocimiento de voz.**

Para comenzar a utilizar el reconocedor de voz y operar el sistema robótico móvil, el usuario tendrá que indicarle los comandos de operación mediante la voz. La interfaz está diseñada para estar escuchando en cada instante de tiempo. Cuando el usuario termine de hablar, en ese momento se ejecutan las bibliotecas configuradas de HTK y los archivos generados en la fase de adaptación, para proporcionar una salida, es decir, una frase de operación. Una vez tomada la frase, se analiza y se confirma la acción al usuario. Por otro lado, se reproducirá artificialmente la frase obtenida para confirmar con el usuario la orden dada. Suponiendo que el usuario responda que sea correcto, se procederá analizar la frase reconocida y se obtendrá un código de la frase de operación. El código obtenido se envía a la Interfaz cliente-servidor para ejecutar el comando de movimiento. Además, la interfaz estará en estado de espera para recibir la confirmación que se esté ejecutando el movimiento indicado. En el caso que exista algún objeto u obstáculo que impida la ejecución o culminación de la orden, el robot se detendrá notificándole al usuario que la orden no se puede realizar, así como los posibles movimientos que podrían ordenarse alternativamente, a esta fase se le definió como “navegación segura”.

## 2.4 Operación de la plataforma robótica móvil

Los comandos de voz para la operación del sistema robótico móvil se basan en el Modelo de lenguaje. Para poder utilizar el reconocedor de voz se definieron 5 códigos de operación del sistema robótico móvil (ver Tabla 2).

Tarea	Código	Ángulo	Velocidad
Desplazamiento	a	-	0.2, 0.5 y 1.0
Retroceder	r	-	0.2, 0.5 y 1.0
Paro	p	0	0
Girar izquierda	i	30 - 180	0.5
Girar derecha	d	30 - 180	0.5

**Tabla 2. Códigos de operación del sistema robótico móvil.**

Por otro lado, es necesario enviar información entre la interfaz y el robot móvil, de manera que una vez reconocido el código de operación la plataforma móvil supiera que operación debe ejecutar. Cabe mencionar, la plataforma robótica ERA-MOBI (ver Fig. 5) está diseñado para trabajar con una computadora abordo, la cual se conecta de forma inalámbrica. La computadora controla la arquitectura de software Player/Stage, que a su vez permite controlar los distintos componentes de un robot a través de un sencillo sistema de interfaces genéricas. Además, mediante la interfaz cliente-servidor permite comunicarse por medio del protocolo TCP/IP con la interfaz gráfica de usuario enviando señales de control y recibiendo información.



Fig. 5. Plataforma robótica ERA-MOBI.

### 3. Resultados

Inicialmente, el SRAH base fue probado con el mismo corpus de entrenamiento para verificar la exactitud en el modelado de fonemas. La métrica para medir el rendimiento del sistema fue el porcentaje de exactitud en las palabras reconocidas, el cual es obtenido mediante la expresión (1).

$$\%WAcc = \frac{N - D - S - I}{N} \quad (1)$$

Donde N es el número de palabras de referencia (frases correctas), S, D e I, son el número de palabras Sustituidas, Eliminadas e Insertadas en la frase reconocida respectivamente. Los resultados obtenidos con las mismas frases del corpus de entrenamiento muestran un porcentaje de reconocimiento del 98.26% (ver Tabla 3).

% de reconocimiento	N	D	S	I
98.26	9990	42	70	62

Tabla 3. Desempeño del reconocedor con las frases de entrenamiento.

Posteriormente se probó el sistema con nuevos usuarios, es decir, se adaptaron 4 usuarios (2 hombres y 2 mujeres) diferentes a las 10 personas que apoyaron en el corpus de entrenamiento. Para ello los usuarios debieron leer 16 frases de adaptación y así poder utilizar el Sistema de Reconocimiento de Voz. Los resultados obtenidos de los nuevos 4 usuarios adaptados muestran un porcentaje de reconocimiento de 80.06% (ver Tabla 4), además se observa el porcentaje de reconocimiento obtenidos del sistema original, es decir, sin realizar la fase de adaptación. Como se puede observar en este trabajo, se obtuvo mejor rendimiento en la tasa de precisión, utilizando la fase de adaptación.

<b>Usuarios</b>	<b>% de reconocimiento adaptado</b>	<b>% de reconocimiento sin adaptación</b>
Hombre 1	80.73	67.79
Hombre 2	78.25	50.26
Mujer 1	81.91	56.77
Mujer 2	79.35	67.42
Total	80.06	60.56

**Tabla 4. Desempeño del reconocedor con las frase de adaptación.**

Finalmente, en las pruebas realizadas para la operación del sistema robótico móvil, se utilizaron los mismos 10 usuarios con que se entrenó el sistema. Cada usuario leyó 23

frases de operación y se obtuvieron un total de 115 muestras de voz, para hombres y mujeres. En este caso, ya no se utilizó el porcentaje de palabras reconocidas como medida de desempeño, si no el número de frases reconocidas de manera correcta. Con ello, se obtuvieron los resultados mostrados en la Tabla 5, se puede observar un porcentaje total promedio de reconocimiento del 94.39%.

<b>Usuarios</b>	<b>fallas/total</b>	<b>% de reconocimiento</b>
Hombres	7/115	93.99
Mujeres	6/115	94.79

**Tabla 5. Desempeño del reconocedor con las frase de adaptación.**

#### **4. Conclusiones**

En este trabajo se presentó el diseño y desarrollo modular de una interfaz humano-máquina capaz de reconocer comandos de voz para operar una plataforma robótica móvil mediante una navegación segura, es decir, evitando colisiones. El diseño modular del sistema permite avanzar hacia la implantación del SRAH sobre una silla de ruedas robotizada para asistir a personas con discapacidades motrices.

Para la construcción del SRAH es de gran ayuda contar con herramientas y librerías como las que provee el HTK, sin embargo esto por sí solo no es suficiente para lograr buenos resultados de reconocimiento, ya que se debe conocer cómo utilizarlas de manera correcta y esto no es una tarea fácil. No obstante, en el presente trabajo se lograron resultados de reconocimiento bastante aceptables tanto para usuarios del corpus como para usuarios nuevos.

Como trabajo a futuro se pretende aumentar el corpus de entrenamiento, para desarrollar un sistema de reconocimiento automático del habla más robusto y aumentar el modelo de lenguaje, con un número mayor de palabras a reconocer. Posteriormente, el sistema será incluido al actual proyecto del Cuerpo Académico de Ingeniería en Computación de



la Universidad del Istmo, el cual consiste en el desarrollo de una silla de ruedas autónoma con brazo robótico para asistir a personas con capacidades disminuidas, ésta además de desplazarse y evadir obstáculos, tendrá un brazo robótico que permitirá tomar objetos del ambiente. Por lo que se pretende utilizar la interfaz de comunicación para poder controlar mediante voz las funcionalidades u operaciones de dicha silla de ruedas, así como las operaciones del brazo robótico.

## 5. Referencias

- [1] I.H. Villamil Espinosa. Aplicaciones en Reconocimiento de Voz utilizando HTK. Tesis de Maestría en Electrónica. Pontificia Universidad Javeriana. Santa Fe de Bogotá, DC. 2005.
- [2] A.B. Caballero Pedrera. Reconocimiento automático de habla con adaptación al género y al locutor. Tesis de Ingeniería en Telecomunicaciones. Universidad Carlos III de Madrid. Departamento de Teoría de la Señal y Comunicaciones. Leganés. 2010.
- [3] U. M. García García. Módulo de reconocimiento de voz a texto independiente de locutor para sistemas de diálogo. Tesis de Ingeniería en Electrónica. Pontificia Universidad Católica del Perú, 2011.
- [4] L.A. Pineda. El proyecto DIME y el robot conversacional Golem: Una experiencia multidisciplinaria entre la computación y la lingüística. Universidad Nacional Autónoma de México. UNAM. 2008. Recuperado de: <http://turing.iimas.unam.mx/~luis/DIME/> , julio de 2014.
- [5] V.M.E. García. Análisis de actos del habla con el esquema dime-damsl: Modelación de diálogos prácticos en transacciones para la interacción humano-computadora. Ph.D. dissertation, Universidad Nacional Autónoma de México, 2013.

- [6] L.A. Pindea. Grupo Golem: RoboCup@Home 2013. Proceedings of Robocup 2013. Vol, pp. 2013.
- [7] H.H Avilés-Arriaga, L.E. Sucar, E.F. Morales, B.A. Vargas, J. Sánchez & E. Corona. Markovito: A flexible and general service robot. In Design and Control of Intelligent Robotic Systems (pp.401-423). Springer Berlin Heidelberg.
- [8] L.E. Sucar, E.F. Morales, H.H. Avilés-Arriaga, P. Oropeza & R. Luis. Markovito's Team Description RoboCup@ Home 2014. Instituto Nacional de Astrofísica, Óptica y Electrónica. 2014.
- [9] M. Hawley, P. Enderby, P. Green, S. Brownsell, A. Hatzis, M. Parker, J. Carmichael, S. Cunningham, P. O'Neill, R. Palmer. STARDUST; Speech Training And Recognition for Dysarthric Users of Assistive Technology. In Proceedings of the 7th European Conference for the Advancement of Assistive Technology in Europe. Dublin, Ireland. 2003.
- [10] J.O. Cuátara Priede. Fonética de la ciudad de México. Aportaciones desde las tecnologías del habla. Tesis para obtener el título de Maestro en Lingüística Hispánica. Universidad Nacional Autónoma de México. 2004.
- [11] D. Jurafsky, J.H. Martin. Speech and Language Processing. Second Edition. 2000. Pearson: Prentice.
- [12] S. Young, P. Woodland. The HTK Book (for HTK Version 3.4). Cambridge University Engineering Department. 2006.
- [13] C.J. Leggetter, P.C. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov model. *Computer Speech and Language*, 9(2):171-185, 1995.
- [14] B. Shneiderman. Designing the user interface: strategies for effective human-computer interaction. Addison-Wesley Reading, MA. 4th ed. 2004.

## **6. Autores**

Héctor Gutiérrez Hernández es tesista colaborador del Cuerpo Académico de Ingeniería en Computación de la Universidad del Istmo, campus Tehuantepec. Ha realizado estancias profesionales en las empresas Datateam Consulting y en Metasoftica S. C. También ha participado en el concurso de mini robótica de la Universidad Tecnológica de la Mixteca en 2011 y en el concurso de programación de la Universidad del Golfo en el mismo año. Actualmente está por concluir su proyecto de tesis para obtener el título de Ingeniero en Computación.

M. en C. J. Jesús Arellano Pimentel obtuvo su título de Maestría en Ciencias en Ingeniería Eléctrica con especialidad en Sistemas Computacionales por la Universidad Michoacana de San Nicolás de Hidalgo en 2005. Actualmente se desempeña como Profesor-Investigador adscrito a la carrera de Ingeniería en Computación de la Universidad del Istmo, campus Tehuantepec. Además es RCA del CA de Ingeniería en Computación de la Universidad del Istmo.

M. en C. Daniel Pacheco Bautista obtuvo su título de Maestría en Ciencias con Especialidad en Electrónica por el Instituto Nacional de Astrofísica, Óptica y Electrónica en 2003. Actualmente es estudiante del programa de doctorado en Ingeniería Biomédica, en la Universidad Popular Autónoma del Estado de Puebla, y es Profesor Investigador en la Universidad del Istmo, campus Tehuantepec. Además, es miembro activo del CA de Ingeniería en Computación de la Universidad del Istmo.