



SISTEMA DE NAVEGACIÓN POR VOZ PARA ROBOTS MÓVILES AUTÓNOMOS

M. J. L. BOADA¹, R. BARBER², V. EGIDO³, M. A. SALICHS²

¹Universidad Carlos III de Madrid, Departamento de Ingeniería Mecánica

²Universidad Carlos III de Madrid, Departamento de Ingeniería de Sistemas y Automática
Avda. de la Universidad 30, 28911-Leganes, Madrid, España

³Universidad Europea de Madrid, Departamento de Arquitectura de Computadores y Automática
C/ Tajo s/n, Urbanización El Bosque, 28670-Villaviciosa de Odon, Madrid, España

(Recibido 10 de marzo de 2004, para publicación 4 de diciembre de 2005)

Resumen – En la actualidad, está proliferando el uso de la voz como interfaz para la interacción con robots móviles, en lugar de los métodos tradicionales, como son el teclado, el ratón, botones, etc. Esto es debido a que, para los humanos, el lenguaje verbal es su medio de comunicación más natural. Éste es el ámbito en el que se sitúa el presente trabajo, que tiene como objetivo dotar al robot de la capacidad de recibir órdenes a través de la voz y de ejecutar las acciones requeridas. En esta primera aproximación, al robot se le indican órdenes de movimiento para guiarlo en un entorno típico de interiores formado por pasillos y habitaciones. Todo ello se hace incorporando e integrando nuevas funcionalidades y habilidades a la arquitectura de control de robots móviles AD. Para su implementación ha sido necesario trabajar en los dos niveles de dicha arquitectura. En el nivel Automático se han implementado un sensor virtual que interacciona con el micrófono, una habilidad que reconoce comandos a partir de las palabras recogidas del micrófono y una acción refleja que dota al robot de un mecanismo rápido y seguro para que se detenga. En el nivel Deliberativo, se ha desarrollado un navegador que gestiona y supervisa la correcta ejecución de las habilidades del nivel Automático.

Palabras clave – Robots autónomos móviles, arquitectura de control de robots, navegación por voz.

1. INTRODUCCIÓN

La aparición de nuevas generaciones de robots cada día más integrados en nuestra sociedad y en nuestros entornos más cercanos, como pueden ser robots personales [1,2], mascotas o robots asistenciales, introduce la necesidad de formular nuevos paradigmas de interacción entre seres humanos y robots. Por ello, se tiende a interfases todavía más cercanas a los humanos, donde se intenta que el robot sea capaz de comunicarse con el humano en el lenguaje más natural y propio de los humanos: el oral. Así, han aparecido recientemente trabajos que tienen en cuenta el lenguaje oral como medio de interacción [3,4], y en particular, como asistencia a la navegación [5,6]. Todos ellos son consecuencia del importante avance en el campo del procesado de la voz [7,8].

En los últimos años, la robótica ha ido evolucionando, tratando de prestar servicios útiles a los hombres e integrándose cada vez más en su entorno. La interacción entre el hombre y los robots será tanto mayor cuanto más se parezcan los robots a los seres humanos tanto en el aspecto físico como psíquico.

En cuanto al aspecto físico es cada día mayor la cantidad de trabajos que aparecen en el diseño y control de robots humanoides. Así lo muestran los trabajos de Laschi y Okumura [9,10], donde se muestran robots con forma humana que intentan reproducir movimientos humanos como andar o arrodillarse.

En el aspecto psíquico se intenta copiar las formas de razonar, navegar y aprender; también se intenta reproducir la estructura emocional humana e incluso se diseñan arquitecturas completas que intentan implemmentar los niveles de abstracción mental y afectiva del ser humano.

Es lógico pensar que estas formas de interacción entre los robots y seres humanos, que acercan las capacidades operativas del robot al humano, deben acercarse a las capacidades de comunicación de los humanos. Las primeras interfases hombre-robot han sido a través de comandos suministrados por un ope-

rador al robot a través de las computadoras. Posteriormente estas interfases han ido evolucionando a interfases gráficas donde el humano podía interactuar con el robot mediante un ratón manipulando elementos gráficos en una pantalla. Otra forma de interacción clásica ha sido la teleoperación, donde el humano da órdenes a un robot manipulando una reproducción a escala del mismo.

Actualmente empiezan a considerarse otras formas de interacción como son la voz, la interacción visual y la táctil. Así, mediante la voz, el robot podrá recibir órdenes e información o comunicar su situación. De igual manera, el robot podrá captar indicaciones de dirección o de cualquier otro tipo mediante una cámara o podrá a su vez proporcionarnos información a través de una pantalla o mediante unas luces. En cuanto a la comunicación táctil, el robot podrá captar tanto una orden dada presionando una de sus partes o incluso percibir una simple caricia.

El presente artículo tiene como objetivo el dotar al robot móvil autónomo de la capacidad de recibir órdenes a través de la voz y de ejecutar las acciones requeridas. En esta primera aproximación, al robot se le indican órdenes de movimiento para guiarlo en un entorno típico de interiores formado por pasillos y habitaciones. Para ello, se va a desarrollar un sistema que permita navegar utilizando comandos de voz, de forma segura y efectiva, en un robot móvil. El sistema propuesto, se engloba dentro de la arquitectura de control AD (Automatic-Deliberative Architecture) desarrollada por Barber [11]. En esta arquitectura se pretende dotar al robot de habilidades y comportamientos que imiten al ser humano en la forma de razonar y de actuar, y de esta manera, lograr tener una interacción más intuitiva con éste. La arquitectura AD consta de dos niveles: el nivel Deliberativo, donde se agrupan las capacidades de razonamiento y toma de decisiones, y el nivel Automático, que engloba las capacidades de percibir el entorno y actuar sobre él. Para que el robot sea capaz de procesar la voz y ejecutar los comandos de navegación mediante la voz, es necesario implementar y añadir diferentes módulos en la arquitectura de control AD, tales como sensores virtuales, habilidades automáticas y deliberativas, y acciones reflejas.

2. ARQUITECTURA AD

De acuerdo con las teorías de la psicología moderna [12], existen dos mecanismos mediante los que el hombre procesa la información percibida de su entorno: los procesos automáticos y los controlados. Por procesos controlados se entiende aquellos procesos que requieren de la capacidad de razonamiento o decisión. Los procesos automáticos apenas requieren de intervención de la atención. En el ser humano cabe, por lo tanto, diferenciar entre dos niveles de actividad mental: nivel Deliberativo y el nivel Automático. Basándose en estas ideas, en la arquitectura AD se establecen sólo dos niveles (Fig. 1): Deliberativo y Automático. El nivel Deliberativo esta asociado con procesos reflexivos y el nivel Automático con procesos automáticos.

Ambos niveles presentan una característica común: están formados por habilidades. Las habilidades son las diferentes capacidades para razonar o llevar a cabo una acción. Estas habilidades son activadas por órdenes de ejecución producidas por otras habilidades o por un secuenciador, devolviendo datos y eventos a las habilidades o secuenciadores que las hayan activado. Esas habilidades son la base de la arquitectura AD.

La comunicación entre los niveles Deliberativo y Automático es bidireccional (Fig. 1). Ambos niveles se comunican a través de la memoria a corto plazo, en la que se encuentra la información acerca del estado del robot, y de dos flujos de información:

- Flujo de eventos: Las habilidades situadas en el nivel Automático notifican a las habilidades del nivel Deliberativo los eventos que generan.
- Flujo de órdenes de ejecución: Las habilidades automáticas reciben órdenes de ejecución procedentes del nivel Deliberativo. Las habilidades deliberativas deciden qué habilidades del nivel Automático tienen que ser activadas o desactivadas en cada momento.

2.1. Nivel Deliberativo

En el nivel Deliberativo se encuentran los módulos que requieren razonamiento o capacidad de decisión. Estos módulos son las habilidades deliberativas. Las habilidades deliberativas no producen respues-

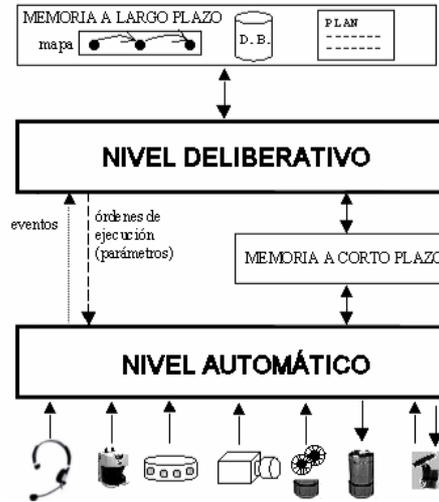


Fig. 1. Niveles de la arquitectura AD.

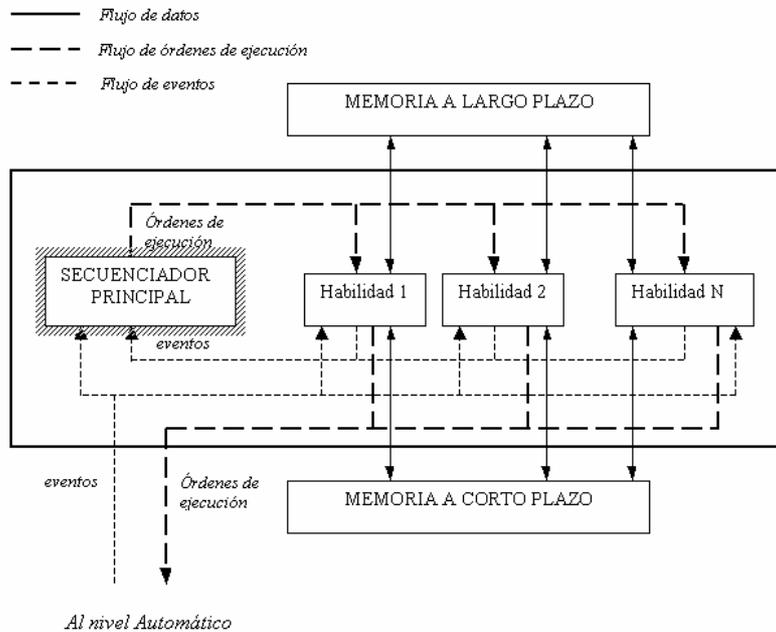


Fig. 2. Nivel Deliberativo.

tas inmediatas ya que necesitan tiempo para tomar decisiones en función de la información que disponen. Dichas habilidades son activadas por el Secuenciador Principal (ver Fig. 2), que es el encargado de su coordinación para conseguir que el robot ejecute correctamente la tarea encomendada. A diferencia del nivel Automático, las actividades del nivel Deliberativo se llevan a cabo secuencialmente y no es posible realizar más de una actividad deliberativa al mismo tiempo, tal y como ocurre en los seres humanos (los seres humanos solamente pueden razonar sobre una cosa al mismo tiempo). La secuencia principal de ejecución necesaria para el cumplimiento de una misión viene establecida a priori y va a marcar el comportamiento del robot y cómo éste va a proceder en las diversas situaciones.

Ejemplos de implementación de habilidades deliberativas son los planificadores, navegadores y los módulos que realizan el modelado del entorno [14,15]. Entre estas habilidades se encuentra la habilidad *Navegador con Voz* encargada de activar las habilidades del nivel Automático, que se describirá más adelante.

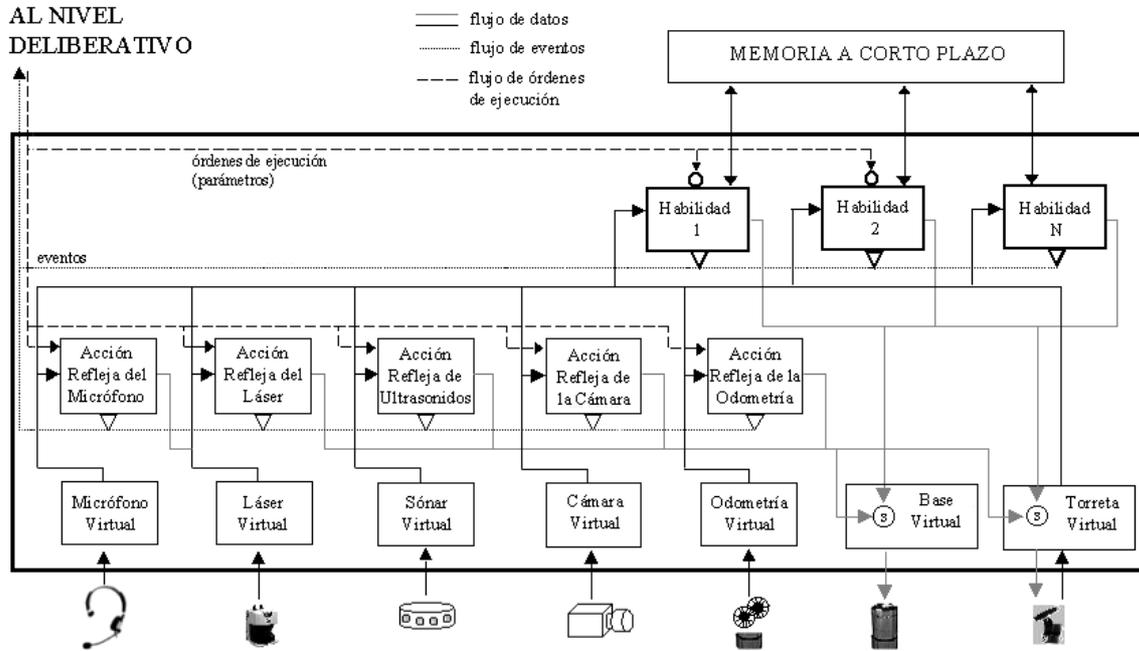


Fig. 3. Nivel Automático.

2.2. Nivel Automático

El Nivel Automático se encarga del control a bajo nivel de los dispositivos de los que dispone el robot. Este nivel permite al robot tener la capacidad de reacción necesaria para responder rápidamente a cambios producidos en su entorno [16]. La Fig. 3 muestra los elementos que constituyen este nivel:

- Sensores y actuadores virtuales: El Nivel Automático se comunica con el hardware del robot a través de los sensores virtuales y los actuadores virtuales.
- Acciones reflejas: Son las respuestas involuntarias y prioritarias a los estímulos.
- Habilidades automáticas: Son las capacidades sensoriales y motoras del sistema y son la base del concepto de comportamiento.

El nivel Automático debe cumplir los siguientes requisitos:

- Los elementos que constituyen este nivel deben ser capaces de ejecutarse en paralelo y de interactuar unos con otros.
- Debe permitir añadir nuevos elementos sin afectar el funcionamiento de los otros, de manera que la arquitectura sea ampliable y dinámica.
- Debe permitir la generación de habilidades complejas a partir de las existentes. Una habilidad puede ser utilizada por varias habilidades complejas, lo que aporta flexibilidad a la arquitectura.

2.2.1. Sensores y actuadores virtuales

Los sensores y actuadores virtuales son los módulos que se comunican con el hardware del robot. Los módulos que necesiten acceder a datos proporcionados por los sensores o enviar comandos a los actuadores lo harán a través de ellos.

Los sensores virtuales proporcionan a las habilidades automáticas y a las acciones reflejas información leída de los sensores físicos del robot. Estos sensores informan sobre el entorno del robot (láser, sónar, micrófono, etc.), o su estado interno (odometría, etc.). Los sensores virtuales almacenan la información obtenida de los sensores físicos en un formato adecuado para ser usado por otros elementos.

Los actuadores virtuales envían comandos de movimiento procedentes de las habilidades automáticas o desde las acciones reflejas a los actuadores físicos del robot. Los actuadores virtuales permiten a las acciones reflejas tener prioridad para enviar comandos de movimiento. Cuando se produce una acción refleja, ésta envía comandos a los correspondientes actuadores inhibiendo el envío de comandos procedentes de las habilidades automáticas. Cuando la acción refleja desaparece, las habilidades automáticas pueden enviar de nuevo comandos a los actuadores. Un actuador no sólo recibe comandos de movimiento, también puede recibir comandos para informar sobre su estado interno: velocidad con la que se mueve o posición en la que se encuentra.

2.2.2. Acciones Reflejas

Un reflejo es la forma más simple del comportamiento animal. Es una respuesta automática, rápida e involuntaria a un determinado estímulo imprevisto, interno o externo, y su duración depende directamente de la duración e intensidad del estímulo que la causa [17]. Las acciones reflejas son acciones rápidas y prioritarias y no implican la actuación del cerebro, aunque éste puede recibir información sobre la localización y la naturaleza del estímulo. Una vez que el cerebro es consciente de que se ha producido una acción refleja, éste puede alterarla o inhibirla. De acuerdo con estas ideas, cada acción refleja está asociada con un sistema sensorial. Esto significa que habrá una acción refleja asociada al sónar, otra asociada al sistema de visión, al láser etc. En la mayoría de los casos, las acciones reflejas se producen cuando la lectura de los sensores exceden un cierto umbral, como por ejemplo, si la lectura del sónar indica que el robot está situado a una distancia peligrosa de la pared, la acción refleja hará que le robot se aleje de ella a tanto mayor velocidad cuanto menor sea dicha distancia. La acción refleja está siempre leyendo los datos procedentes de los sensores. Si el dato que percibe excede un límite, la acción refleja entra en un estado reflejo. En este caso, la acción refleja se conecta con los sensores virtuales para mandarle comandos de movimiento, inhibiendo el envío de comandos procedentes de las habilidades automáticas. Al mismo tiempo, el nivel Deliberativo recibe notificación de la acción refleja. En ese momento, el nivel Deliberativo puede alterar o inhibir la acción refleja a través de las órdenes de ejecución. La acción refleja termina cuando el estímulo reflejo termina o cuando la acción refleja es inhibida por el nivel Deliberativo.

2.2.3. Habilidades automáticas

Las habilidades automáticas se definen como la capacidad de procesamiento de la información sensorial y/o la ejecución de acciones sobre los elementos motores del robot. Las habilidades automáticas se pueden clasificar en función de los tipos de datos que obtienen como resultados y de cómo interactúan con los dispositivos hardware del robot, es decir, sensores y actuadores. Según este tipo de clasificación la habilidades pueden ser perceptivas y sensorimotoras.

Las habilidades perceptivas son aquellas que interpretan la información procedente de los sensores u otras habilidades perceptivas o sensorimotoras. Estas habilidades no generan comandos de movimiento para los actuadores sino que los resultados del procesamiento se almacenan en sus objetos de datos correspondientes a los que pueden conectarse y acceder otras habilidades perceptivas o sensorimotoras para ser utilizados como datos de entrada por éstas.

Las habilidades sensorimotoras son aquellas que producen comandos de movimiento en función de la información percibida de los sensores, habilidades perceptivas o sensorimotoras. Para enviar órdenes de movimiento, la habilidad sensorimotora se conecta al objeto de datos del actuador virtual correspondiente y actualiza sus valores.

Las habilidades perceptivas interpretan la información recibida desde los sensores, de las habilidades sensorimotoras o de otras habilidades perceptivas. Las habilidades sensorimotoras perciben información desde los sensores, habilidades perceptivas o de otras habilidades sensorimotoras y la procesan con el fin de ejecutar una acción sobre los actuadores.

Todas las habilidades automáticas presentan las características siguientes:

- Pueden ser activadas por habilidades situadas en el mismo nivel o en el nivel Deliberativo. Una habilidad sólo puede desactivar habilidades que ella misma ha activado previamente.
- Deben almacenar los resultados de forma tal que puedan ser utilizados por otras habilidades.

- Pueden generar eventos. Solamente notificarán el evento producido a aquellas habilidades que lo hayan solicitado previamente.

Las habilidades automáticas sencillas de la arquitectura AD constituyen el substrato para la generación de habilidades complejas. Estas habilidades a su vez se pueden combinar para dar lugar a habilidades más complejas. El carácter modular de las habilidades permite utilizarlas fácilmente para construir jerarquías de habilidades con mayores niveles de abstracción. Las habilidades automáticas no están organizadas *a priori*. Las habilidades se configuran dinámicamente en función de la tarea a realizar por el robot y de las condiciones del entorno. Con esto se consigue una mayor flexibilidad en los comportamientos del robot.

2.3. Memorias de la arquitectura AD

El conocimiento permite a un robot aprender a partir de lo experimentado anteriormente, predecir situaciones que se puedan producir, y razonar o tomar decisiones de las tareas a ejecutar o de cómo llevarlas a cabo. Para ello, el robot tiene que ser capaz de adquirir y retener dicho conocimiento. En la memoria no sólo se almacena la información percibida directamente por los sensores sino también la información obtenida de su procesamiento y del aprendizaje.

En función de la durabilidad del conocimiento almacenado, y teniendo en cuenta el modelo de memoria desarrollado por Atkinson y Shiffrin [18], en la arquitectura AD se distinguen tres tipos de memoria: memoria sensorial, memoria a corto plazo y memoria a largo plazo. La información almacenada en los dos primeros tipos de memoria es transitoria y está relacionada con el procesamiento llevado a cabo por las habilidades automáticas mientras que la información almacenada en la memoria a largo plazo es persistente o más estable a largo plazo. Dicha información es utilizada por las habilidades del nivel Deliberativo para razonar o tomar decisiones.

La memoria sensorial se caracteriza porque en ella se almacena la información percibida directamente de los sensores, tiene una alta capacidad de almacenamiento y corta duración. En la arquitectura AD dicha memoria está formada por los objetos de datos contenidos en cada uno de los sensores virtuales, debido a que en ellos es donde se guarda la información procedente directamente de los sensores. A medida que el robot se mueve por el entorno adquiere nueva información olvidando la percibida anteriormente. En función de la importancia que dicha información tenga para el robot, ésta puede ser almacenada en la memoria a corto plazo. Esta información se guarda en una forma más abstracta como resultado del procesamiento de la información sensorial. Es decir, sólo se almacena la información que es útil para que el robot ejecute un comportamiento determinado. En la arquitectura AD, parte de esta información está contenida en las habilidades automáticas y deliberativas y otra es almacenada en una zona común para ambos niveles.

En la memoria a largo plazo se almacena el conocimiento que debe permanecer en el tiempo. Este conocimiento puede venir del aprendizaje, del procesamiento de la información almacenada en la memoria a corto plazo o puede estar dada *a priori*. El robot usa este conocimiento para razonar o para tomar decisiones.

3. IMPLEMENTACIÓN DE LA NAVEGACIÓN POR VOZ EN LA ARQUITECTURA AD

En este apartado se describe la implementación de los distintos módulos involucrados en la navegación mediante comandos de voz en la arquitectura de control AD para robots móviles autónomos. Dichos módulos son: habilidad deliberativa *Navegador con Voz*, habilidad automática *Reconocimiento de Comandos de Voz*, Micrófono Virtual y acción refleja asociada al Micrófono Virtual. La Fig. 4 muestra el esquema completo de la implementación. Los otros módulos presentes en dicha arquitectura han sido desarrollados en trabajos previos [16,19].

Para la implementación de todos los módulos de la arquitectura AD, se ha utilizado el paquete software de *Mobility*. *Mobility* [20] es un paquete software distribuido, integrado y orientado a objetos para el control de robots individuales y sistemas multirobot. Está formado por:

- Un conjunto de herramientas software.
- Un modelo de objetos para el software del robot.

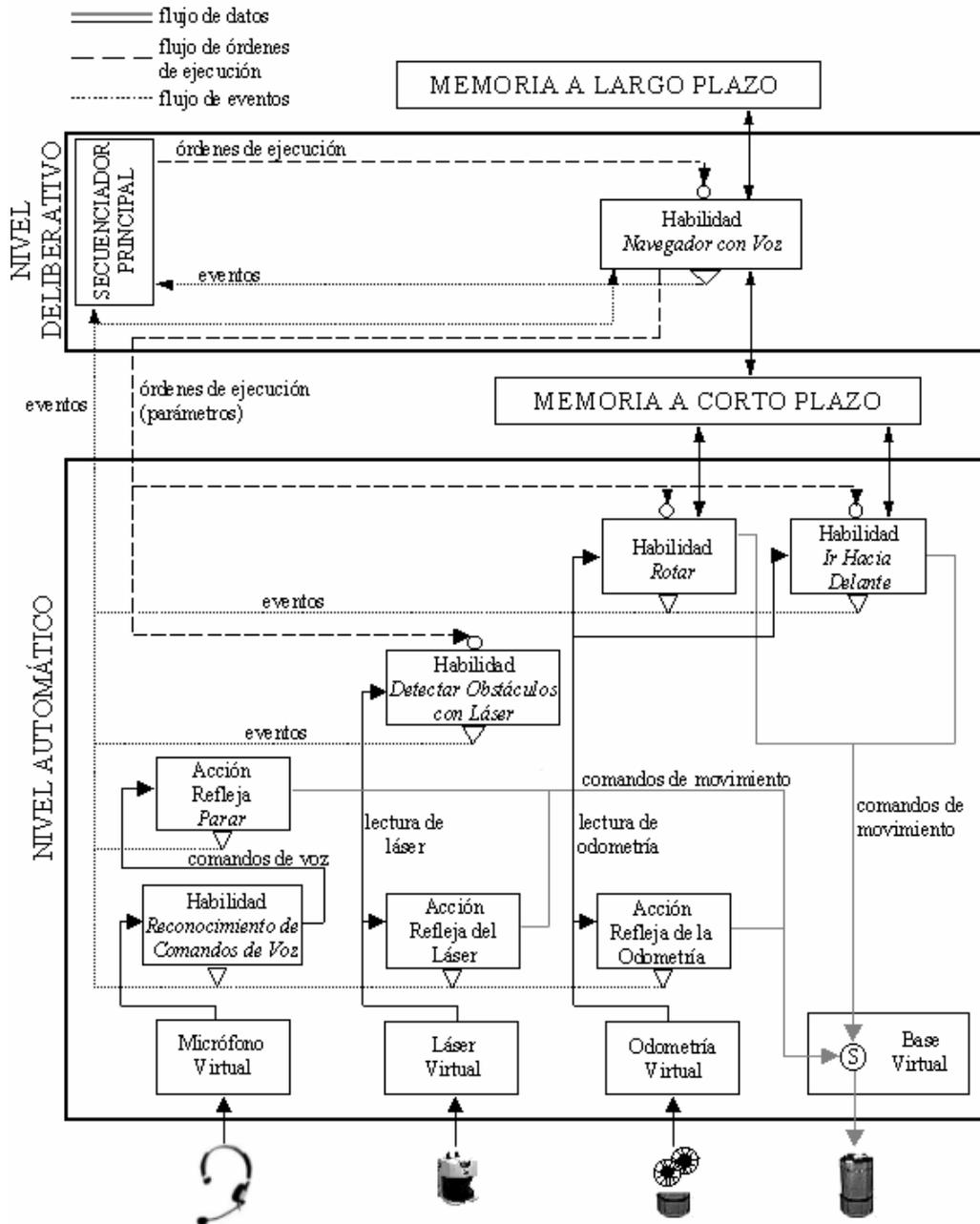


Fig. 4. Implementación de la navegación mediante voz en la arquitectura AD.

- Un conjunto de módulos de control básicos del robot.
- Una librería de clases orientada a objetos para simplificar el desarrollo del código.

La arquitectura de objetos definida en Mobility se basa en objetos organizados jerárquicamente y unidos entre sí mediante el estándar CORBA 2.X, el cual emplea el lenguaje IDL para describir las interfaces entre objetos, donde cada uno de ellos representa de forma abstracta sensores, actuadores, etc. Una de las ventajas más importantes de esta arquitectura es que cada módulo es una parte independiente, por lo que cualquier modificación que se haga sobre él no afecta al resto de la arquitectura. Las ventajas importantes que Mobility aporta son:

- Procesado paralelo y distribuido para el control de robots.



Fig. 5. Estructura del Micrófono Virtual.

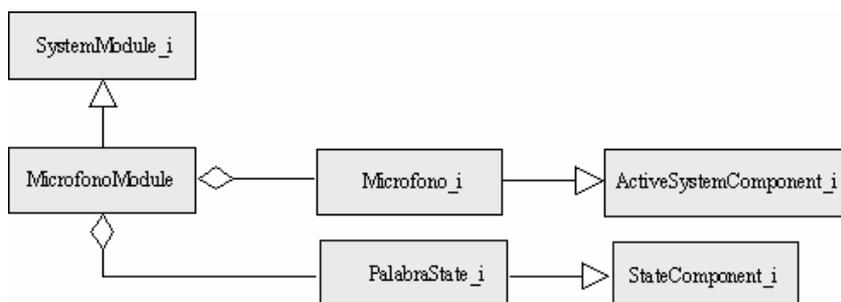


Fig. 6. Jerarquía parcial de clases del Micrófono Virtual.

- Alta reutilización del software.
- Integración de código.
- Desarrollo de sistemas multirobot.

Todos los módulos que forman la arquitectura AD son módulos cliente-servidor. Cada módulo puede contener objetos distintos. Estos objetos son unidades de software separadas con su propia identidad, interfaces y estados. Para la comunicación entre módulos se utiliza CORBA.

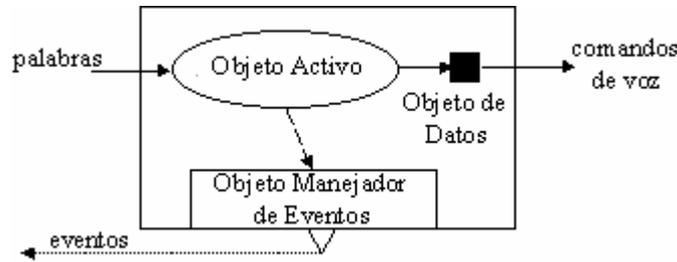
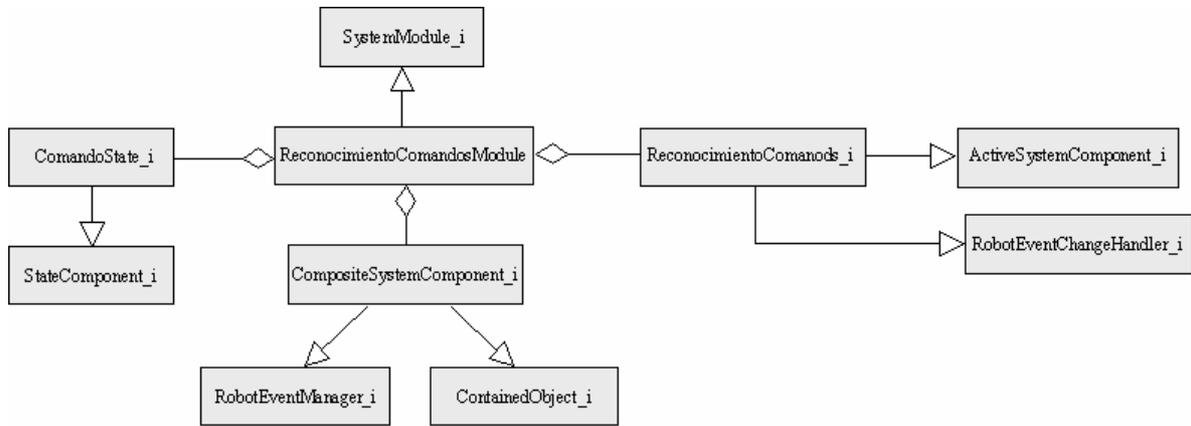
CORBA (*Common Object Request Broker Architecture*) [21] es una arquitectura independiente de la plataforma y del lenguaje de programación que establece un conjunto estandarizado de normas para la comunicación de componentes software distribuidos. CORBA está desarrollada y mantenida por el OMG, Grupo de Gestión de Objetos (*Object Management Group*) que es un consorcio que agrupa tanto a empresas de la industria informática como a los consumidores de la misma.

3.1. Micrófono Virtual.

El Micrófono Virtual es un sensor virtual [16], por lo que se ha desarrollado como un módulo servidor debido a que proporciona recursos a otros módulos. El Micrófono Virtual contiene un objeto activo y un objeto de datos tal y como se muestra en la Fig. 5. El objeto activo es el que tiene su propio hilo de control y, en este caso, se conecta al puerto en el que se encuentra el micrófono. El objeto activo lee continuamente los datos proporcionados por dicho sensor y, cada vez que detecta sonido, es convertido en palabras y almacenado en el objeto de datos del servidor para poder ser utilizado por otros módulos. En este caso, hará uso de esta información la habilidad automática *Reconocimiento de Comandos*.

En la Fig. 6 se representa la jerarquía parcial de clases que constituyen el servidor del micrófono. El módulo del micrófono, denominado *MicrophoneModule*, hereda de la clase *SystemModule_i*. Esta clase permite contener a otros componentes en el módulo y que se puedan comunicar entre ellos. Los objetos que contiene este módulo son:

- *Microphone_i*: Es un objeto activo que lee continuamente las lecturas proporcionadas por el micrófono a través de *CVoiceControl*. Se encarga, también, de enviar las palabras de voz obtenidas, al objeto de datos del módulo. La clase *ActiveSystemComponent* encapsula todos los mecanismos de creación de hilos.

Fig. 7. Estructura de la habilidad *Reconocimiento de Comandos*.Fig. 8. Estructura de la habilidad *Reconocimiento de Comandos*.

- *PalabraState_i*: En este objeto se almacenan las palabras detectadas. La actualización de este objeto de datos es dinámica y asíncrona. El objeto *StateComponent_i* actualiza dinámicamente y soporta múltiples observadores de esta actualización. Este objeto también puede contener una sub-representación de estados (diferentes vistas o niveles de detalle). Todos los objetos que heredan de *StateComponent_i* tienen interfaces similares.

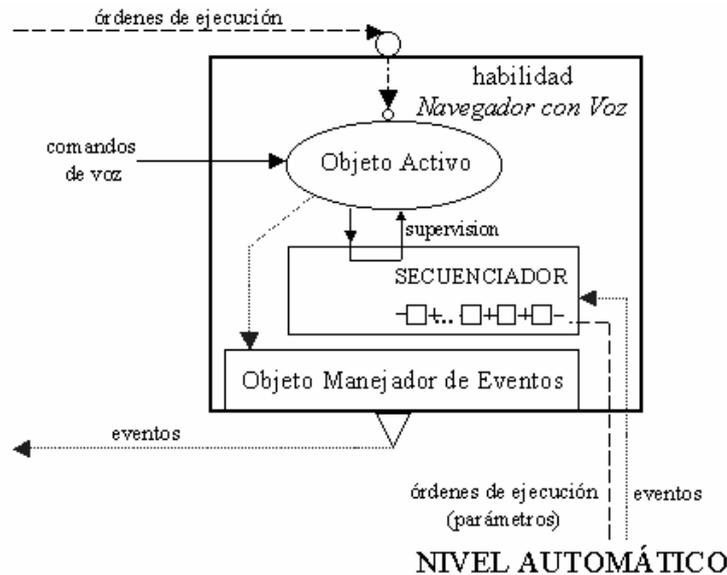
3.2. Habilidad Reconocimiento de Comandos

La habilidad automática *Reconocimiento de Comandos* es una habilidad perceptiva [16]. Dicha habilidad se encarga de recoger las palabras almacenadas en la memoria a corto plazo por el sensor virtual y de extraer los comandos aprendidos previamente. A su vez, las almacena en la memoria a corto plazo para que la habilidad deliberativa *Navegador con Voz* tenga acceso a dichos comandos y ordene y supervise su ejecución. Los comandos reconocidos en esta implementación son: AVANZA, GIRA DERECHA, GIRA IZQUIERDA y PARA.

La habilidad contiene un objeto activo, un manejador de eventos y un objeto de datos, tal y como se muestra en la Fig. 7. El objeto activo es el encargado de reconocer los comandos aprendidos y de notificarlos a la habilidad *Navegador con Voz* a través del manejador de eventos. Los comandos de voz son almacenados en un objeto de datos, estando a disposición de otras habilidades.

La Fig. 8 muestra la jerarquía parcial de las clases que constituyen dicha habilidad. El módulo *ReconocimientoComandosModule* contiene los objetos siguientes:

- *ReconocimientoComandos_i*: Es el objeto activo. Hereda de las clases *ActiveSystemComponent_i*, para la creación y gestión del hilo de control, y de *RobotEventManager_i*. Esta última clase contiene los métodos para que el objeto:
 - Se conecte al objeto manejador de eventos del que quiere recibir notificación.

Fig. 9. Estructura del *Navegador con Voz*.

- Reciba notificación del objeto manejador de eventos correspondiente cuando se produzca dicho evento.
- *RobotEventManagerObj_i*: Es el objeto manejador de eventos. Este objeto guarda en una lista la dirección CORBA de las habilidades que se registran en él y el evento del que quieren recibir notificación. Cada vez que se produce un evento, el objeto activo de la habilidad se lo comunica a su manejador de eventos y éste a todas las habilidades que se han registrado en él y que quieren recibir notificación de ese evento.
- *ComandoState_i*: Es el objeto de datos que contiene los comandos identificados por la habilidad.

3.3. Habilidad *Navegador con Voz*

La habilidad *Navegador con voz* es una habilidad deliberativa ya que gestiona y supervisa habilidades del nivel Automático. Cada vez que recibe notificación de la habilidad *Reconocimiento de comandos* de que un nuevo comando ha sido reconocido, se encarga de la activación y desactivación de la habilidad automática correspondiente. Si la orden no puede realizarse, el navegador tiene que decidir que hacer. Por ejemplo, se le indica al robot que avance y el robot tiene delante una pared, el navegador toma la decisión de parar el robot, comunicando al operador humano su decisión y quedando a la espera de un nuevo comando.

En la Fig. 9 se muestra la estructura de la habilidad deliberativa *Navegador con voz*. La habilidad consta de un objeto activo, encargado de gestionar el hilo de control de la habilidad, un secuenciador que activa y desactiva las habilidades del Nivel Automático, un manejador de eventos, que notifica eventos al secuenciador principal y un supervisor, incluido en el objeto activo y que supervisa la correcta ejecución de la tarea de navegación.

3.4. Acción refleja *PARAR*

Para dotar al robot de un mecanismo rápido y eficiente en su detención, la orden PARA ha sido asociada a una acción refleja [16]. La Fig. 10 muestra la estructura de la acción refleja implementada.

La acción refleja *Parar* contiene un objeto activo que comprueba continuamente las palabras almacenadas en el sensor virtual. Cuando la orden PARA es detectada, el objeto activo entra en el estado *reflejo*. En ese momento, la acción refleja se conecta al actuador virtual encargado del movimiento del robot. El actuador virtual otorga prioridad a la acción refleja y detiene el robot, deshabilitando las ordenes de ac-

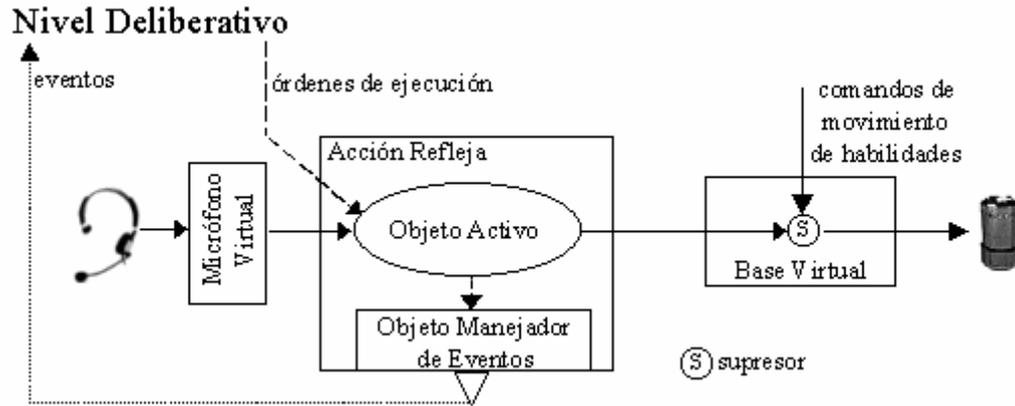
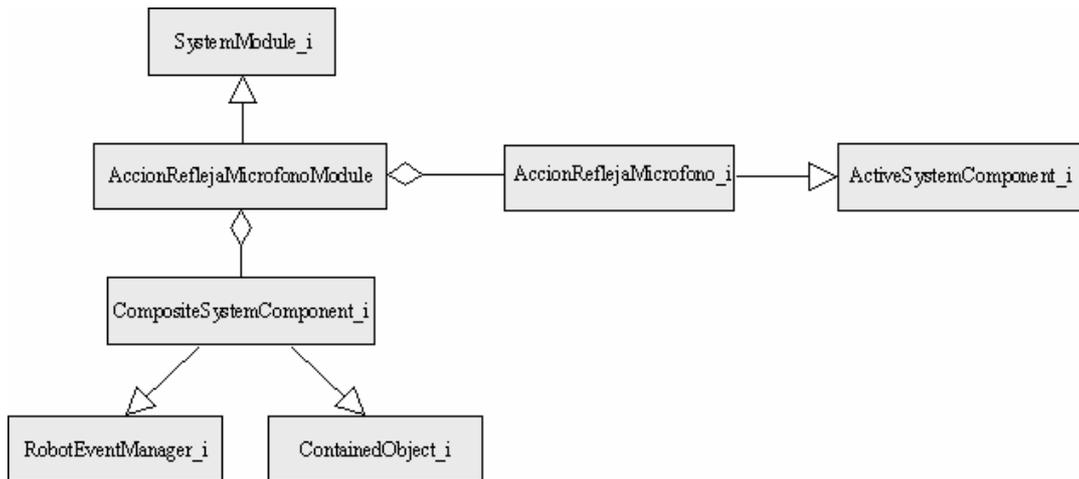
Fig. 10. Estructura de la acción refleja *Parar* asociada al micrófono.

Fig. 11. Jerarquía parcial de clases de la acción refleja asociada al micrófono.

tuación de las otras habilidades automáticas sobre el sistema locomotor del robot. Al mismo tiempo, la acción refleja notifica a la habilidad deliberativa *Navegador con voz* mediante un evento que se ha producido la orden PARA. Una vez que esta habilidad ha recibido la notificación, detiene la tarea de navegación. La acción refleja entra en el estado *no reflejo* cuando el nivel Deliberativo inhibe la acción refleja.

La Fig. 11 muestra la jerarquía parcial de clases que forman la acción refleja asociada al sensor virtual del micrófono.

Junto con la acción refleja *Parar* asociada al micrófono, se ha implementado también la acción refleja asociada al sensor láser. Cuando la lectura del sensor supera un cierto umbral, es decir, indica que el robot se encuentra a una determinada distancia de la pared considerada como peligrosa, la acción refleja hará que el robot se aleje de ella a tanto mayor velocidad cuanto menor sea la distancia. Esta habilidad ha sido implementada para dotar al robot de un cierto grado de autonomía navegando de manera segura y evitando que colisione con algún obstáculo que aparezca de manera imprevista y que el operario no haya visto.

4. ROBOT RWI B21

El B21 de Real World Interface Inc. (RWI) [22] es un robot móvil diseñado como soporte para aplicaciones robóticas de investigación y desarrollo en medios interiores. Este robot, Fig. 12, está dividido en

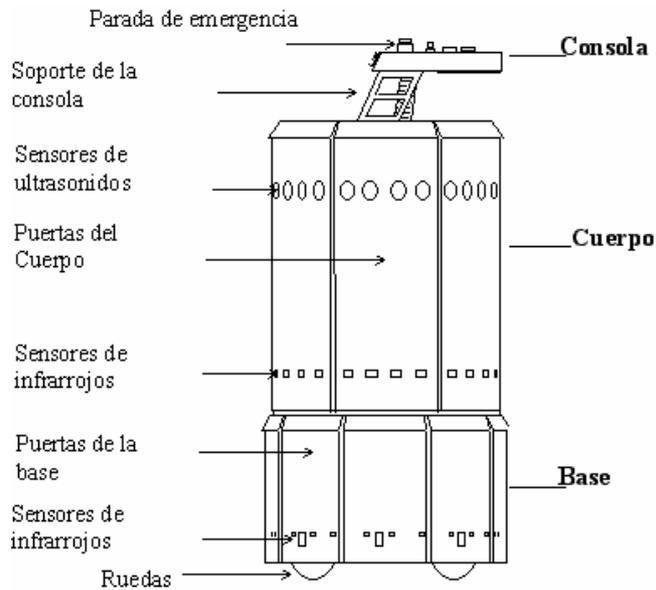


Fig. 12. Robot RWI B21.



Fig. 13. Posición del telémetro láser y sistema de visión en el robot 21.

tres secciones principales: la base, el cuerpo y la consola. El cuerpo y la consola se hallan físicamente unidos, sin posibilidad de separación, mientras que la base es fácilmente separable del cuerpo.

El robot se mueve mediante 4 ruedas que son guiadas de forma sincronizada. La base no rota con las ruedas, al ser éstas dirigidas, sino que permanece siempre con la misma orientación. Es el cuerpo el que se orienta según el ángulo de rotación de las ruedas.

La base consta de los motores, las baterías, las ruedas, los circuitos de freno, y otros componentes de potencia y movimiento. En el cuerpo se hallan los sensores, las CPUs, los equipos de comunicación, el interfaz con la consola, ciertos indicadores, interruptores, y otros equipos de sensado y procesado. La consola se encuentra en la parte superior del cuerpo y contiene la mayor parte de los indicadores e interruptores del B21.

Toda la superficie exterior, tanto del cuerpo como de la base, está cubierta de paneles curvos que funcionan como sensores táctiles, y que pueden ser abiertos fácilmente para acceder al interior del robot. Sobre estos paneles se encuentran los sensores de ultrasonidos (24 sensores), táctiles (24 sensores) y de infrarrojos (24 sensores). Los robots B21 pueden tener uno o dos PCs instalados en el interior del cuerpo. Se puede disponer además de un ordenador portátil situado en la parte superior del cuerpo y que se puede utilizar como interfaz con el robot.

Sobre la base del robot están montados sensores como cámaras o telémetros láser (Fig. 13). El telémetro láser PLS de la casa SICK es un dispositivo capaz de medir distancias a los objetos que se encuentran en su entorno. El telémetro láser realiza un escaneado entre 0° y 180° del entorno proporcionando la magnitud de las distancias a los objetos que se encuentran en el plano de medida. Está colocado en la parte frontal de la consola del robot B21, por lo que el plano de medida está situado aproximadamente a 110 cm del suelo. La conexión con el PC del robot se realiza vía protocolo RS-232 a 9600 baudios.

El sistema de visión artificial empleado está formado por una plataforma orientable, con movimientos *pan* y *tilt* de TRC modelo Zebra, que permite el movimiento de la cámara CCD que va montada sobre ella. El sistema que constituye la torreta tiene incorporado un controlador de motor de tres ejes de la marca ZEBRA, diseñado para el servo control de motores DC con realimentación de posición mediante encoders, acompañado de una unidad para comunicaciones RS-232. Para la adquisición de imágenes se utiliza una cámara de vídeo monocroma XC-75CE que se basa en el empleo de dispositivos de carga acoplada CCD como elementos sensores.

5. RESULTADOS EXPERIMENTALES

Para hacer llegar al robot las órdenes de movimiento o parada, se ha utilizado un micrófono sin hilos. El dispositivo finalmente elegido fue el Shure Serie T, que consta de un micrófono, un emisor y un receptor, y está conectado a la tarjeta de sonido del PC.

Las pruebas se han realizado utilizando cuatro comandos básicos: AVANZA, PARA, GIRA DERECHA y GIRA IZQUIERDA. Se utilizaron varios micrófonos y varios interlocutores. El reconocimiento de comandos se realizó con la ayuda del programa CvoiceControl.

Para mostrar el funcionamiento del sistema, se presentan dos misiones de guiado del robot a través de la voz. La primera consiste en recorrer un pasillo y entrar una habitación. La segunda misión consiste en navegar por el interior de una habitación con obstáculos. En ambas misiones, se toman los datos de odometría de los sensores del robot.

En la misión “Recorrer pasillo y entrar en sala 1”, se parte con el robot en el pasillo, punto A (Fig. 14), y ya orientado aproximadamente en el sentido del pasillo. En B se le dice que gire a la derecha y avance. En C se ordena dos veces “gira derecha”, ya que el ángulo necesario para cruzar la puerta es mayor de 90° . En D la orientación se rectifica girando a la izquierda y parando rápidamente. Después se le ordena al robot avanzar hasta E y que gire a la izquierda. A continuación el robot avanza hasta F, donde gira a la izquierda para intentar ajustar la dirección del robot a la puerta de la sala 2. En la Fig. 15 se muestra el robot durante la misión.

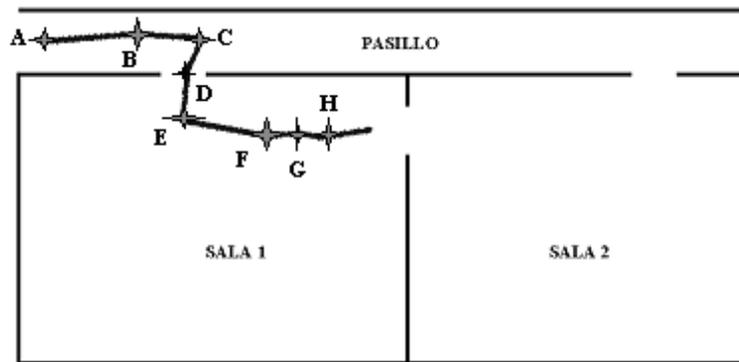


Fig. 14. Odometría de misión recorrer pasillo y entrar en sala 1.



Fig. 15. Robot durante la misión 1.

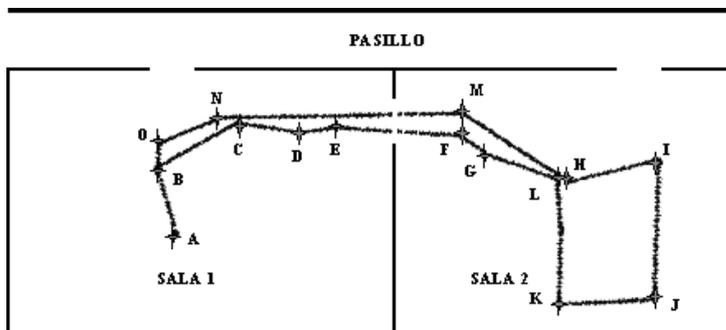


Fig. 16. Odometría de misión navegación interior.



Fig. 17. Robot durante la misión 2.

En la misión de navegación interior, se parte con el robot en la Sala 1, accede a la Sala 2 y rodea e unas mesas, esquivando los obstáculos que hay a su paso. Como se puede observar en la Fig. 16 (C,D,E), para conseguir la orientación precisa para pasar por la puerta, es necesario normalmente girar varias veces ya que cuando se pide parar el giro existen retardos en la ejecución de la orden y así como inercia del robot, que hace que la parada no sea inmediata. En la Fig. 17 se muestra al robot durante esta misión.

Tabla 1. Resultados de éxito obtenidos en el reconocimiento de comandos de voz en diferentes situaciones.

		COMANDOS			
sujeto	micrófono	AVANZA	GIRA IZQUIERDA	GIRA DERECHA	PARA
varón 1	inalámbrico	95%	55%	55%	95%
varón 1	inalámbrico	65%	70%	65%	100%
varón 1	normal	80%	55%	85%	90%
varón 2	inalámbrico	70%	75%	70%	85%
mujer	inalámbrico	85%	95%	60%	95%

Durante las pruebas se almacenaron estadísticas sobre el reconocimiento de los comandos de voz en las situaciones reales (ver tabla 1). Los resultados obtenidos fueron los siguientes:

1. La orden PARA es bastante robusta y está en el 92,5% de media. Esto es muy importante ya que se necesita operar con un sistema que pueda parar cuando sea necesario, por motivos de seguridad. Además es la más rápida en ejecutarse, ya que está implementada como una acción refleja.
2. La orden PARA no tiene falsos positivos, es decir, no se reconocen otras ordenes distintas a la de parar. Esto sería especialmente crítico en el caso de que se reconociera AVANZA, en vez de PARA ya que esto podría hacer que el robot colisionase.
3. Con las órdenes de giro, se obtiene un porcentaje por encima del 60% de media, lo que indica que si la primera orden no fue reconocida, se reconocerá a la segunda. También habría que resaltar que tampoco en las ordenes de giro se reconoce un falso positivo de AVANZA, aunque en varias ocasiones se reconoce como orden el sentido inverso. Esto no es crítico, ya que el robot solo gira sobre su base, con lo que no existe riesgo de colisión.
4. También se observa que con el micrófono con cable, se obtienen mejores tasas de reconocimiento que con el micrófono inalámbrico.
5. Aunque diferentes interlocutores producen diferentes porcentajes, estos se encuentran por encima del 60 %.

Con estos datos, podemos afirmar que el reconocimiento es aceptable y que el robot puede ser gobernado por distintos interlocutores, realizando una navegación segura.

6. CONCLUSIONES

En el presente trabajo se ha implementado un sistema de navegación con voz para interactuar con el robot. Para ello ha sido necesario desarrollar diversos módulos en la arquitectura de control de robots móviles AD. Se han probado diferentes situaciones que influyen en el reconocimiento de comandos en situaciones de operación reales, obteniéndose unos resultados aceptables. El comando PARA tiene un porcentaje de reconocimiento alto y es rápida por estar implementada como una acción refleja, permitiendo una navegación segura.

Para que la interacción sea lo mas completa posible, los futuros desarrollos se orientan a conseguir que el robot sea capaz de interpretar frases, ideas e incluso emociones más allá de lo puramente explícito, como puede ser el caso de que un usuario le mande ir de un sitio a otro, y le comunique el camino a seguir con cierta vaguedad.

AGRADECIMIENTOS

Los autores agradecen la ayuda concedida por el Ministerio de Ciencia y Tecnología de España para el desarrollo de este trabajo de investigación mediante el proyecto DPI2002-00188. Así mismo agradecen las aportaciones realizadas por Diego López en la elaboración de los resultados experimentales.

REFERENCIAS

- [1] Kubota, N.; Nojima, Y.; Baba, N.; Kojima, F. and Fukuda, T. “Evolving pet robot with Emotional Model”. Proceedings of the 2000 Congress on Evolutionary computation (2000).
- [2] Kawamura, K.; Pack, R.T.; Bishay, M. and Iskarous, M. “Design philosophy for service robots”. Robotics and Autonomous Systems, vol.18, no.1-2, p. 109-16 (1996).
- [3] Mumolo, E.; Nolich, M. and Vercelli, G. “Algorithms for acoustic localization based on microphone array in service robotics”. Robotics and Autonomous Systems, vol.42, no.2., p. 69-88 (2003).
- [4] Rodriguez-Losada, D.; Matia, F.; Galan, R. and Jimenez, A. “Blacky, an interactive mobile robot at a trade fair”. Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292), vol.4, p. 3930-5 (2002).
- [5] Pires, G. and Nunes, U. “A wheelchair steered through voice commands and assisted by a reactive fuzzy-logic controller”. Intelligent and Robotic Systems: Theory and Applications, vol.34, p. 301-14 (2002).
- [6] Kubik, T. and Sugisaka, M. “Rule based robot navigation system working in an indoor environment”. Proceedings of the 14th International Conference on Systems Science, vol.2, p. 212-19 (2001).
- [7] Agaram, K., Keckler, S.W. and Burger, D. “A characterization of speech recognition on modern computer systems”. Proceedings of the Fourth WWC-4 IEEE, p. 45-53. (2001).
- [8] Qian-Rong G. and Shibata, T. “Speaker and text independent language identification using predictive error histogram vectors”. Inter. Conf. on Acoustics, Speech and Signal Processing. V1, p. 36-9 (2003).
- [9] Laschi E., E., C. Dario and P. Guglielmelli. “Humanoids and personal robots: design and experiments”. Robotic Systems. 18:673-90 (2001).
- [10] Okumura, T., Y. Shimizu, M. Furuta and Tawara T. “Design and construction of a series of compact humanoid robots and development of biped walk control strategies”. Robotics and Autonomous Systems, 37:81-100 (2001).
- [11] Barber, R; Salichs, M.A. “A new human based architecture for intelligent autonomous robots”. The Fourth IFAC Symposium on Intelligent Autonomous Vehicles, p. 85-90 (2001).
- [12] Shiffrin, R. M. and Schneider, W. “Controlled and automatic human information processing: Perceptual learning, automatic attending and general theory”. Psychological Review, p 127-190 (1997).
- [13] Atkinson, R. C. and Shiffrin, R. M. “The psychology of learning and motivation”, volume 2, chapter Human Memory: A Proposed System and Its Control Processes, p 89-195. New York: Academic Press (1968).
- [14] Egado V., Barber R., Boada M. J. L. and Salichs M. A. “Self-generation by a mobile robot of topological maps of corridors”. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*, p 2662-2667 (2002).
- [15] Egado V., Barber R., Boada M. J. L. and Salichs M. A. “A human orientated planner for topological navigation”. Para ser publicado en The 5th IFAC Symposium on Intelligent Autonomous Vehicles (2004).
- [16] Boada, M. J. L., Barber, R. and Salichs, M. A. “Visual approach skill for a mobile robot using learning and fusion of simple skills”. Robotics and Autonomous Systems, Vol 38, Issues 3-4, p157-170. (2002).
- [17] Xu, H. and Brussel, H. V. “A behavior-based blackboard architecture for reactive and efficient task execution of an autonomous robot”. Robotics and Autonomous Systems, 22:115-132 (1997).
- [18] Shiffrin, R.M., Schneider, W. “Controlled and automatic human information processing: II Perceptual learning, automatic attending and a general theory”. Psychological Review 84, p 127-190. (1997).
- [19] Boada, M.J.L., Rodriguez, F.J., Barber, R., Salichs, M.A. “A control System Based on Reactive Skills for Autonomous Mobile Robots”. The 11th International Conference on Advanced Robotics. (2003).
- [20] “Robot Integration Software. User’s Guide”. Real World Interface. A División of IS Robotics, INC (1999).

- [21] Siegel, J, “CORBA Fundamentals and programming”, Wiley Computer Publishing (1996).
- [22] “B21 robot system manual”. Real Word Interface, Inc (1996).

VOICE NAVIGATION SYSTEM FOR MOBILE AUTONOMOUS ROBOTS

Abstract – Nowadays, traditional methods to interact with mobile robots such as keyboard, mouse, push-buttons, etc. are being replaced with the voice. This is due to the fact that the oral language is the most natural media for human beings. This is the field in which the present paper is located. Its main goal is to provide the robot with the skill to get commands via voice and to execute the request actions. In this first approach, movement commands are given to guide the robot through a typical indoor environment. For its implementation the two levels of the AD robot control architecture have been used. In the Automatic level a virtual sensor that interacts with a microphone, a skill that recognizes commands from the virtual sensor and a reflex action that provides the robot a fast and safe mechanism to stop have been implemented. In the Deliberative level, a navigator that supervises and manages the correct execution of skills from the Automatic level has been developed.

Keywords – Autonomous mobile robot, robot control architecture, voice navigation.

