Interacción adaptada al contexto en una interfaz de diálogos orales para entornos inteligentes

Germán Montoro, Pablo A. Haya y Xavier Alamán

Departamento de Ingeniería Informática
Universidad Autónoma de Madrid

{German.Montoro, Pablo.Haya, Xavier.Alaman}@uam.es

Resumen

En esta comunicación presentamos los procesos de interpretación y generación de una interfaz de diálogos orales para entornos inteligentes. La interfaz se crea de forma automática adaptándose a cada entorno y la interpretación y generación varía dependiendo del entorno y su contexto. Estos procesos se basan en una estructura de diálogos en árbol. Varios módulos procesan la estructura en árbol y la información de contexto del entorno para producir diálogos específicos para el estado actual del entorno. Los diálogos poseen características de clarificación, recuperación de errores, resolución de la anáfora, etc. La interfaz se ha implementado en un laboratorio con un entorno inteligente real.

1.- Introducción

Los entornos inteligentes han aparecido como un nuevo campo de investigación dentro del área de interfaces de usuario. Estos entornos interaccionan con los usuarios de forma natural y les ayudan en sus tareas cotidianas. Los ordenadores y dispositivos computaciones quedan ocultos a los usuarios y éstos pueden obtener los servicios del sistema, por ejemplo, mediante interfaces orales en lenguaje natural sensibles al contexto.

La aparición de estos entornos hace necesario construir nuevas interfaces que permitan interaccionar con los usuarios de forma natural. Estas interfaces tienen que ser capaces de entablar conversaciones relativas al entorno, sus elementos y los servicios que puede proporcionar a sus usuarios. Las interfaces de diálogos orales se han de adaptar a estos sistemas, de modo que puedan hacer frente a los nuevos retos que proporcionan los entornos inteligentes.

En esta comunicación presentamos una interfaz de diálogos orales para entornos inteligentes que se adapta al dominio de cada entorno. Los diálogos se crean de forma automática y permiten interaccionar con el entorno y controlar sus dispositivos mediante interacción oral en lenguaje natural. La comunicación se centra en los procesos de interpretación y generación y sólo explica de forma breve cómo el sistema construye de forma automática los diálogos para un entorno dado.

Para llevar a cabo nuestra investigación hemos construido un entorno inteligente real. Este entorno está formado por un laboratorio amueblado como una sala de estar y en el que se ha dispuesto un conjunto de dispositivos. Entre ellos se encuentran luces y controles de iluminación, un mecanismo de apertura de la puerta de entrada, un detector de presencia, tarjetas de acceso personales, altavoces, micrófonos, un sintonizador de radio, cámaras IP, etc.

El resto de la comunicación se organiza como sigue: la siguiente sección describe muy brevemente cómo se representa el entorno y la interfaz de diálogos para su creación automática; a continuación se explican los procesos de interacción, interpretación y generación y, por último, se presentan las conclusiones y el trabajo futuro.

2.- Representación del entorno y los diálogos

En esta sección se muestra de forma breve cómo se representa la información del entorno y los diálogos que permite la creación automática de la interfaz de diálogos orales. Una explicación detallada de este proceso se puede encontrar en [1].

La representación del entorno se realiza en un documento XML. Al inicio, el sistema lee la información de este documento XML y, además de otros módulos cuya descripción está del alcance de esta comunicación, construye de forma automática:

- Una pizarra [2], que trabaja como una capa de interacción entre el mundo físico y la interfaz de diálogos orales (además de otras aplicaciones e interfaces).
- Una interfaz de diálogos orales que, empleando la pizarra, funciona como una capa de interacción entre los usuarios y el entorno.

La pizarra almacena una representación de múltiples características del entorno. Esta capa de pizarra aísla a la aplicación del entorno real. Los detalles de las entidades del mundo físico y las aplicaciones se ocultan a los clientes [3].

Las entidades de la pizarra se asocian a un tipo de entidad. Todas las entidades del mismo tipo heredan las mismas propiedades generales. Algunas de estas propiedades se corresponden con partes lingüísticas que se emplean para crear la interfaz de diálogos orales. Esta información, entre otros elementos, se compone de:

- Parte verbal (VP). Describe las acciones que se pueden llevar a cabo con la entidad.
- Parte objeto (OP). Establece los nombres que puede recibir.

- Parte de ubicación (LP). Describe su situación física dentro del entorno.
- Parte de objeto indirecto (IOP). Especifica a quién va dirigida la acción que se realiza.
- Parte modal (MODALP). Indica el modo en que se debe realizar la acción.
- Parte cuantificadora (QP). Define un valor o cantidad que se aplica sobre la acción que se realiza.
- Parte modificadora (MP). Añade información calificativa a alguna de las partes anteriores.

La estructura del diálogo se basa en un árbol lingüístico. Cada conjunto de partes lingüísticas se transforma en una ruta del árbol, con un nodo para cada parte.

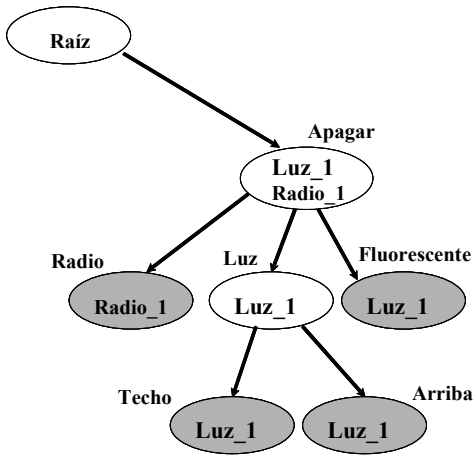

Figura 1: Árbol lingüístico para las entidades luz_1 y radio_1

Como ejemplo, supongamos que la entidad *luz_1* tiene los siguientes conjuntos de partes lingüísticas para la acción de apagar:

{"VP : apagar", "OP : luz", "LP : techo arriba"}
{"VP : apagar", "OP : fluorescente"}

Esto es, el primero de ellos se compone de la parte verbal apagar, la parte objeto *luz* y las partes de ubicación *techo* y *arriba,* que son sinónimos. El segundo está formado por la parte verbal *apagar* y la parte objeto *fluorescente*. Estos conjuntos de partes se combinan con una gramática adecuada, lo que permite establecer múltiples posibilidades para realizar la acción apagar con la entidad *luz_1*.

En este caso, la parte verbal *apagar* está al mismo nivel en ambos conjuntos de partes, de modo que sólo se crea un único nodo "apagar". De este mismo nodo colgarán los nodos "luz" y "fluorescente", mientras que del nodo "luz" colgarán los nodos "techo" y "arriba". En cada nodo se guardará la entidad que tiene asociada, en este caso se almacena *luz_1* en todos los nodos.

A continuación se añade una entidad llamada *radio_1* con el siguiente conjunto de partes:

{"VP : apagar", "OP : radio"}

El sistema añade el nombre de la entidad *radio_1* al nodo "luz" y a continuación crea un nuevo nodo "radio" que contiene la entidad *radio_1* y que estará al mismo nivel que os nodos "luz" y "fluorescente".

Empezando con un árbol vacío, el sistema crearía de forma automática el árbol lingüístico representado en la figura 1. Los nodos sombreados corresponden con nodos de acción, mientras que los blancos son nodos intermedios.

3.- Interpretación y generación

Una vez creada la interfaz el sistema y el usuario pueden empezar a establecer diálogos sobre el entorno. La interfaz está controlada por un supervisor de los diálogos que se ocupa de recibir la oración del reconocedor de voz, interpretarla y generar un resultado (ya sea mediante una respuesta oral o una acción en el entorno).

Inicialmente, el sistema está dormido y no reconoce ninguna frase de los usuarios. Si un usuario quiere iniciar un diálogo debe despertar al sistema pronunciando la palabra *Odisea*. En ese momento, siguiendo las ideas definidas por [4], el sistema considera que el objetivo de todas las oraciones pronunciadas es completar una acción. Si el usuario no dice nada en los siguientes siete segundos (o siete segundos después de la última interacción) el sistema vuelve al estado dormido.

3.1. Interpretación

Cuando el sistema recibe una oración del reconocedor el supervisor la envía al módulo de procesamiento de oraciones (MPO). El MPO busca coincidencias entre la oración y los hijos de la raíz del árbol lingüístico. Si se produce alguna coincidencia, el MPO desciende en el árbol al nodo donde ésta se produjo y repite el proceso con sus hijos. Este proceso continúa hasta que el MPO alcanza un nodo de acción (lo que significa que la oración se ha interpretado completamente y el sistema puede ejecutar una acción) o hasta que no se producen más coincidencias (por lo que el sistema necesita clarificación).

3.1.1. Ejecución de una acción

Si el MPO alcanza un nodo de acción envía la información del nodo al módulo del procesamiento del nodo (MPN). El MPN obtiene de la ruta comprendida entre la raíz y el nodo de acción al nodo que contiene la parte verbal (esto es, la acción solicitada por el usuario). A continuación, ordena la ejecución de la acción definida en la parte verbal por parte del método asociado al nodo de acción alcanzado. Esta ejecución suele implicar un cambio físico en el entorno aunque también puede suponer una respuesta por parte del sistema.

Por ejemplo, dado el árbol lingüístico representado en la figura 1, supongamos que un usuario pronuncia: *Por favor, podrías apagar la luz del techo*. El MPO recibe la oración y busca coincidencias en el árbol lingüístico. *Apagar* coincide con la parte verbal del nodo "apagar", *luz* coincide con el hijo de "apagar" y *techo* coincide con el hijo de "luz". El nodo "techo" es un nodo de acción. El MPN toma el control y obtiene que la acción solicitada por el usuario es *apagar*. Dado que este nodo está asociado a la entidad *luz_1* ejecuta el método de acción de esta entidad, considerando que la acción requerida por el usuario es *apagar*. De este modo, la ejecución

de este método de acción puede tener dos efectos: apagar la *luz_1* o informar al usuario de que esta entidad ya estaba apagada. En cualquier caso, el MPO considera que la acción se ha completado y regresa a su estado inicial, esperando nuevas oraciones.

3.1.2. Clarificación

Si, después de que el usuario pronuncie una oración, el MPO no alcanza un nodo de acción es necesario realizar clarificación de la oración. El MPN envía la información del nodo del MPO. Este, de nuevo, obtiene la parte verbal (la acción solicitada por el usuario) pero no ejecuta ninguna acción. En esta ocasión el MPO envía la información del nodo al módulo de procesamiento del árbol (MPA). El MPA visita todos los hijos del nodo actual, construyendo una oración de respuesta que solicite más información al usuario. Esta oración se basa en los nodos visitados y el contexto del entorno representado en la pizarra.

La búsqueda del MPA se basa en una búsqueda recursiva en profundidad del árbol. Su procedimiento es:

- Visita el primer hijo del nodo y comprueba si la acción solicitada por el usuario es diferente al estado físico actual de alguna de las entidades asociadas al nodo hijo, empleando la pizarra. En ese caso, almacena el nombre de las entidades con un estado diferente y su nivel en el árbol. Además, obtiene la palabra asociada al nodo, con el fin de construir la oración de respuesta.
- A continuación continúa de forma recursiva con el primer hijo del nodo procesado, siguiendo los mismos pasos.
- Este proceso recursivo con el primer hijo de cada nodo se repite hasta que el MPA alcanza un nodo de acción o hasta que el nodo hijo no tiene ninguna entidad con un estado en el entorno diferente al solicitado. Si el MPA alcanza un nodo de acción, y la entidad asociada a este nodo tiene un estado diferente al solicitado, incrementa en uno el número de acciones que se pueden ofrecer al usuario y almacena el nombre de la entidad que produce la acción. En cualquier caso, el MPA asciende al nodo padre y continúa la inspección recursiva del árbol con el siguiente hijo del nodo.
- Cuando el MPA visita un nodo que no corresponde con el primer hijo, se asegura en primer lugar de que las entidades que contiene no son las mismas que las entidades de otros nodos ya procesados en el mismo nivel. Esto hace posible que los sinónimos se puedan representar en el árbol, pero que sólo se considere el adecuado en los procesos de interpretación y generación.
- Este algoritmo se repite hasta que el MPA ha inspeccionado todos los nodos hijos de la raíz y, en caso de ser necesario, sus hijos subsecuentes.

El hecho de considerar sólo aquellos nodos que tengan entidades con un estado diferente al solicitado por el usuario y el de procesar sólo el primer sinónimo coincidente optimiza el proceso de búsqueda en el árbol. El MPA sólo tiene que seguir las rutas a los nodos de acción que se pueden procesar de acuerdo con el estado actual del entorno, obviando todas las demás.

Una vez que el MPA ha visitado todos los nodos correspondientes, posee una oración de respuesta completa (ver más adelante la sección sobre generación) y el número de acciones que se pueden ofrecer al usuario. El MPO recibe esta información y, dependiendo del número de acciones posibles, se comporta como sigue:

- Si el número de acciones es igual a cero, el MPO informa al usuario de que, en las condiciones actuales, no existe ningún elemento del entorno que permita realizar la acción requerida.
- Si el número de acciones es igual a uno, el MPO ejecuta directamente el método de acción asociado a la única entidad que se ha encontrado que pueda realizar la acción solicitada. Con esto, se reducen el número de turnos y se asiste al reconocedor de voz, empleando el contexto actual del entorno como una de las posibles fuentes de información que mejoran la interpretación y comprensión [5].
- Si el número de acciones es de dos o tres, el MPO pronuncia la oración de clarificación generada por el MPA durante la búsqueda en el árbol lingüístico. Esta oración informa sobre todas las posibles entidades que pueden realizar la acción requerida, guiando de este modo al usuario en el diálogo [6].
- Si el número de acciones es superior a tres, el MPO no ofrece al usuario todas las posibilidades de forma exhaustiva. En su lugar, pronuncia una oración de clarificación basada en la oración generada por el MPA pero simplificada. Con esto, el sistema no utiliza una oración con demasiadas opciones, que puede resultar tediosa y difícil de recordar, aunque todavía guía al usuario [7].

Como se ha podido ver, la interpretación varía dependiendo del contexto actual del entorno. La misma oración reconocida puede conducir a interpretaciones diferentes en contextos distintos. En contextos dispares el MPO puede ejecutar una acción o considerar que el usuario quiere interaccionar con alguna entidad del entorno.

Como ejemplo, supongamos dos casos diferentes para el entorno representado en la figura 1. Ambos comparten el mismo escenario, donde *luz_1* y *radio_1* se encuentran encendidas:

- En el primer caso el usuario pronuncia *Apaga la luz*, de modo que el MPO para en el nodo "luz". Entonces el MPN obtiene que la acción solicitada es *apagar* y el MPA inicia el proceso de clarificación. En primer lugar desciende al nodo "techo" y obtiene que el estado en el entorno de *luz_1* (encendido) es diferente al solicitado (apagar), por lo que procesa el nodo. A continuación, añade la entidad *luz_1* a la lista de entidades visitadas bajo el nodo "luz", agrega la palabra *techo* a la oración de respuesta y, dado que se trata de un nodo de acción, incrementa el número de acciones que se pueden ofrecer y almacena el nombre de la entidad de acción *luz_1*. A continuación, regresa al nodo "luz" e inspecciona el nodo "arriba". En ese punto comprueba si las entidades de ese nodo se encuentran en la lista de entidades procesadas bajo el nodo "luz" (esto es, si es un sinónimo de un nodo procesado previamente). Como el nodo "arriba" sólo contiene la entidad *luz_1* que se encuentra en la lista de nodos procesados para el nodo "luz", omite este nodo (es un sinónimo de uno ya procesado). Finalmente, regresa al

nodo original, lo que concluye la búsqueda en el árbol. El MPA devuelve la oración de respuesta, el número de acciones posibles y la lista de las entidades de acción al MPO. Este comprueba que sólo existe una posible acción y la ejecuta, esto es, apaga la luz del techo.

- En el segundo caso, el usuario pronuncia *Quiero que apagues*, de modo que el MPO para en el nodo "apagar". Nuevamente, después de que el MPN obtiene que la acción solicitada es *apagar* el MPA inicia el proceso de clarificación. En primer lugar obtiene que el estado de *radio_1* es diferente al estado solicitad y que es un nodo de acción. El MPA añade la entidad *radio_1* a la lista de entidades procesadas bajo el nodo "apagar", incrementa el número de acciones posibles, adjunta la palabra *radio* a la oración de respuesta y añade la entidad *radio_1* a la lista de entidades de acción. A continuación, regresa al nodo "apagar" y acto seguido continúa con al nodo "luz". Una vez allí, realiza los mismos pasos explicados en el caso anterior. Finalmente, regresa nuevamente al nodo "apagar" y comprueba el nodo "fluorescente". Este nodo contiene la misma entidad que la de un nodo visitado previamente al mismo nivel (es un sinónimo de un nodo ya procesado) por lo que no se considera. El MPO recibe la oración de respuesta, el número de acciones posibles y la lista de entidades de acción. Como el número de acciones posibles es igual a dos pronuncia la oración de clarificación generada: *¿Prefieres apagar la radio o la luz del techo?* Se puede comprobar que, como se ha explicado anteriormente, los sinónimos se omiten en la respuesta de clarificación (la oración sólo se refiere a la luz del techo y no al fluorescente).

Estas mismas oraciones pueden producir interpretaciones distintas en un contexto diferente. Supongamos un escenario donde *radio_1* está encendida y *luz_1* apagada. Si el usuario repite las dos oraciones previas:

- Para la oración *Apaga la luz* el MPO informará en este caso al usuario de que todas las luces se encuentran apagadas.
- Para la oración *Quiero que apagues* el MPO apagará la *radio_1*, dado que es la única entidad de acción con un estado diferente al solicitado por el usuario.

Otra situación posible se produce cuando existe más de una entidad del mismo tipo en el entorno (por ejemplo, dos o más luces). En este caso para la oración *Apaga la luz* se pueden dar tres posibles circunstancias:

- Si todas las luces están apagadas el MPO informará al usuario sobre este respecto.
- Si sólo hay una luz encendida el MPO apagará esa luz.
- Si hay más de una luz encendida el MPO pronunciará una pregunta de clarificación. Esta oración de clarificación sólo se referirá a aquellas luces que se encuentran encendidas, omitiendo a aquellas que ya se encuentren apagadas.

Como se ha mostrado con estos ejemplos, la interpretación varía dependiendo de las entidades presentes en el entorno y de su estado en cada momento. Se pueden añadir o eliminar entidades al entorno, las entidades presentes pueden cambiar su estado o puede haber múltiples entidades del mismo tipo. La interfaz de diálogos se adapta a estas situaciones de forma automática, modificando su estructura y comportamiento.

Finalmente, después de que el sistema pronuncie una oración de clarificación el proceso de interpretación del MPO sufre una modificación. Tras esta circunstancia, el MPO empieza buscando coincidencias desde los nodos del árbol lingüístico donde se paró en la última interacción. Sólo en el caso de no obtener ningún resultado satisfactorio para esta búsqueda el MPO iniciará la misma búsqueda empezando desde la raíz del árbol. De este modo se permite que el usuario pueda continuar con un diálogo iniciado en interacciones previas (tras una respuesta de clarificación) o empezar un nuevo diálogo (obviando la respuesta de clarificación proporcionada por el sistema).

3.1.3. Recuperación de errores

Los procesos de reconocimiento de voz, interpretación y clarificación pueden conducir, en algunos casos, a fallos en el reconocimiento o la interpretación. A su vez, los usuarios pueden no proporcionar toda la información necesaria para procesar completamente la oración. Para recuperarse de algunos de estos problemas, el sistema proporciona algunas características específicas.

Como se ha visto anteriormente, después de la respuesta de clarificación el MPO permite continuar con la ruta de un diálogo previo o iniciar uno nuevo. Esta característica permite recuperarse de errores en la interpretación. Si la respuesta de clarificación proporcionada por el sistema no corresponde con ninguno de los objetivos del usuario éste puede iniciar un nuevo diálogo desde el principio, en lugar de continuar con un diálogo erróneo.

Como una importante característica adicional, el MPO no sólo comprueba la raíz del árbol lingüístico y el nodo donde se detuvo en la última interacción. Si no se producen coincidencias a partir de estos nodos también buscará coincidencias a partir de cada uno de los hijos de estos nodos. Con esto, el sistema se puede recuperar de errores en el reconocimiento de voz o interpretar correctamente oraciones donde el usuario sólo proporcionó parte de la información. Por ejemplo, si para el escenario representado en la figura 1 el reconocedor devuelve la oración *la luz del techo* el MPO no obtendrá ninguna coincidencia para el nodo raíz. Entonces empezará a buscar coincidencias con los hijos del nodo raíz, esto es, a partir del nodo "apagar". De este modo obtiene coincidencias con el nodo "luz" y con su hijo el nodo "techo". Las coincidencias con estos nodos serán sólo válidas en el caso de que la entidad *luz_1* estuviera encendida. El sistema se he recuperado de un error de reconocimiento (o simplemente el usuario no especificó qué acción quería llevar a cabo) y gracias al uso del contexto puede interpretar correctamente [8] que la oración original debería ser *Apagar la luz del techo*.

3.1.3. Resolución de la anáfora

El sistema de diálogos orales permite la resolución de la anáfora pronominal al referirse a la última entidad mencionada.

Para permitir la utilización de la anáfora se modifica de forma automática el árbol de modo que contenga nodos de resolución de la anáfora. El árbol contiene uno de estos nodos por cada

verbo, formado por el verbo y un pronombre. Además, después de que el MPO alcanza un nodo de acción y ejecuta la acción correspondiente almacena en un historial la ruta seguida y la oración pronunciada por el usuario.

Cuando el MPO alcanza un nodo de resolución de la anáfora (el usuario ha utilizado un pronombre para referirse a un elemento) se sitúa en el nodo del árbol que corresponde con el verbo almacenado en el nodo. Una vez allí desciende a través de la ruta especificada en el historial para la última acción hasta que llega a un nodo de acción y ejecuta la acción asociada.

Supongamos que el usuario pronuncia *Podrías encender la radio*. El MPO alcanza el nodo de acción "radio", enciende la radio y almacena en el historial la ruta seguida hasta ese nodo. A continuación el usuario pronuncia *Por favor, me gustaría que la apagaras*. Ahora el MPO alcanza el nodo de resolución de la anáfora "apagarlo" y desde ahí salta al nodo "apagar". Una vez allí, sigue la ruta almacenada en el historial, esto es, desciende al nodo "radio". Como este es un nodo de acción, ejecuta la acción correspondiente, apagando la radio.

3.2. Generación

Como se ha visto anteriormente, el proceso de generación se realiza al mismo tiempo que el de clarificación. La oración de respuesta se forma mediante las palabras de los nodos que tienen entidades con un estado distinto al solicitado.

Inicialmente la oración de respuesta está vacía. Antes del proceso de clarificación, se inicializa con la palabra *Prefieres* y las palabras de la ruta del árbol comprendida entre la raíz y el nodo donde paró el MPO. Las palabras se añaden a la oración de respuesta mediante un módulo de adición de palabras (MAP). El MAP obtiene el número y género de la palabra y adjunta su forma correcta a la oración de respuesta precedida, en caso de ser necesario, por su correspondiente artículo. Tras la inicialización, el MPA adjunta el resto de palabras que obtiene durante el proceso de clarificación. Las palabras se obtienen tal y como se explicó en la sección sobre clarificación. Adicionalmente, si ya se había añadido a la oración una palabra de un nodo al mismo nivel se añade una coma (,) o la palabra *o* para la última de las posibilidades. Además, antes de las partes modificadoras y de ubicación se añade la preposición *de*, teniendo en cuenta si se debe contraer con el artículo *el*, y antes de la parte de objeto indirecto se añade la palabra *a*.

Por ejemplo, si en la figura 1 el MPO para en el nodo "apagar", la oración de clarificación se inicializará con la frase *Prefieres apagar*. A continuación el MPA utiliza el MAP para añadir las palabras *la radio*. Hecho esto, dado que el nodo "luz" está al mismo nivel que el nodo "radio" añade la palabra *o* seguida de las palabras *la luz*. Acto seguido, como *techo* es una parte de ubicación añade la preposición *de* seguida de las palabras *el techo*. La composición final de la respuesta de clarificación es *Prefieres apagar la radio o la luz del techo*.

Como se puede ver, la generación de oraciones también emplea el contexto y se adapta al estado del entorno en cada momento, obteniendo de forma automática oraciones acordes a cada situación dada.

Adicionalmente, la interfaz no sólo emplea las oraciones de respuesta generadas sino que también utiliza señales de audio [9]. Estas señales consisten en sonidos de ambiente que intentan proporcionar información de forma ligera, intentando no molestar a los usuarios en caso de no ser necesario. Estas señales se emplean en dos circunstancias:

- Se produce una señal de audio similar a un bostezo cuando el reconocedor de voz retorna al estado dormido. Si esto ocurrió porque el usuario finalizó su interacción con el entorno, no es necesario distraerle informándole sobre este asunto. En caso contrario, el usuario todavía está prestando atención a la interacción con el entorno, por lo que una señal de audio es suficiente para hacerle saber el nuevo estado del reconocedor de voz.
- Se produce una señal de audio diferente, similar a una interjección, cuando el MPO no obtiene ninguna coincidencia a partir del nodo raíz, sus hijos, nodos previos o sus hijos. Esta señal se repite tras la segunda vez consecutiva que resulta imposible realizar ningún tipo de interpretación y, sólo tras un tercer fallo, el sistema genera una oración solicitando al usuario que cambie el tipo de oración empleada. Se introdujo esta señal al comprobar que, en algunos casos, el reconocedor de voz producía errores de sustitución o inserción, esto es, devolvía una oración diferente a la pronunciada o interpretaba el ruido como una oración del usuario [10]. En el primer caso, resulta más rápido y eficiente reproducir una señal de audio que emitir una oración completa. En el segundo caso, un error del reconocedor de voz no tiene por qué distraer al usuario.

4. Conclusiones

En esta comunicación se ha presentado una interfaz de diálogos orales que se adapta a la configuración de cada entorno inteligente. La adaptación ocurre durante el proceso de creación de la interfaz y en el de interacción con el sistema. En ambos casos la interfaz y su comportamiento varían dependiendo del entorno y su estado.

En un futuro se deben añadir nuevas capacidades a las partes de interpretación y generación. Los diálogos actuales permiten interaccionar con las entidades del entorno pero no preguntar por su estado. Se deben realizar algunas modificaciones para permitir realizar preguntas. La interfaz debe contener una nueva parte interrogativa con un nuevo árbol de preguntas, muy similar al árbol lingüístico explicado. Sólo unas pocas características de los procesos de interpretación y generación han de ser modificadas para permitir el uso de este nuevo árbol.

Actualmente el sistema sólo permite resolver una única acción por turno. Por ejemplo, si el usuario solicita dos acciones diferentes en la misma oración sólo se considera la primera. El MPO se debe adaptar para poder llevar a cabo múltiples acciones simultáneas.

El uso de aproximaciones multimodales puede beneficiar a la interfaz. Se planea incorporar al sistema un nuevo módulo de reconocimiento de caras, de modo que se pueda identificar quién está en el entorno. Esta información podría ser utilizada por varios módulos del sistema, incluyendo la interfaz de diálogos orales, para mejorar su funcionalidad.

Siguiendo con esta vertiente, la sincronización del habla y los gestos de los usuarios puede ayudar a mejorar la interacción [11]. Para esto, se debería construir un nuevo módulo de reconocimiento de gestos. Otra posible modalidad de interacción consiste en mostrar la información en una pantalla, en lugar de informar oralmente. En este caso, el usuario podría elegir entre responder utilizando la voz o seleccionado una opción.

5. Agradecimientos

Este trabajo está financiado por el Ministerio de Educación y Ciencia, número de proyecto TIN2004-03140.

6. Referencias

[1] Montoro, G., Alamán, X. and Haya, P.A. "Interacción con entornos inteligentes mediante diálogos basados en contexto". Congreso de interacción Persona Ordenador (Interacción 2004). Lleida. May 3-7, 2004.

[2] Engelmore, R. And Mogan, T. "Blackboard Systems". Addison-Wesley, 1988.

[3] Salber, D. and Abowd, G.D. "The design and use of a generic context server". In Proceedings of Perceptual User Interfaces (PUI'98), 1998.

[4] Searle, J. "Speech Acts". Cambridge University Press. London, 1969.

[5] Ward, K. and Novick, D.G. "Integrating multiple cues for spoken language understanding". In Proceedings of CHI'95, Denver, May 7-11, 1995.

[6] Yankelovich, N. "How do users know what to say?" ACM Interactions, 3, 6, December, 1996

[7] Marx, M. and Schmandt, C. "MailCall: Message presentation and navigation in a nonvisual environment". In Proceedings of CHI'96 (Vancouver, April 13-18), 1996.

[8] Nagao, K. and Rekimoto J. "Ubiquitous talker: Spoken language interaction with real world objects". In Proceedings of IJCAI-95, Vol. 2, 1284-1290, 1995.

[9] Mynatt, E.D.; Back, M.; Want, R. and Frederick, R. "Audio Aura: Light-weight audio augmented reality". In Proceedings of ACM UIST'97 (Banff, Canada), 211-212, 1997.

[10] Schmandt, C. and Negroponte, N. "Voice communication with computers: conversational systems". Van Nostrand Reinhold, New York. 1994.

[11] Bourguet, M. and Ando, A. "Synchronization of speech and hand gestures during multimodal human-computer interaction". In Proceedings of CHI'98, Los Angeles, April 18-23, 241-242, 1998.