

Agentes virtuales 3D para el control de entornos inteligentes domóticos

Eva Cerezo, Sandra Baldassarri, Elisa Cuartero, Francisco Serón
Dept. de Ingeniería de Sistemas e Informática
Universidad de Zaragoza. Instituto de Investigación en Ingeniería de Aragón (I3A)
{ecerezo, sandra, seron}@unizar.es

Germán Montoro, Pablo A. Haya, Xavier Alamán
Dept de Ingeniería Informática
Universidad Autónoma de Madrid
{german.montoro, pablo.haya, xavier.alaman}@uam.es

Resumen

Se están presentando nuevas formas de interacción en el área de los ambientes inteligentes. La interacción multimodal con actores virtuales está cambiando la manera de comunicarse e interactuar con aplicaciones y con otros usuarios. Se presenta en este trabajo un nuevo interfaz basado en un agente virtual 3D que permite la interacción vía voz con un sistema domótico remoto. Se ha logrado una interacción multimodal y natural, por lo que se ha aumentado la accesibilidad del sistema, ampliando su uso potencial a usuarios con discapacidad o de edad avanzada. Su desarrollo se ha basado en el sistema Maxine, un motor de generación y control de escenarios y caracteres virtuales especialmente orientado a la generación de nuevos interfaces.

Palabras clave: domótica, agentes virtuales 3D, interfaces naturales

1. Introducción

Los entornos inteligentes enfatizan en interfaces más amigables para el usuario, servicios más eficientes, más control por parte del usuario y soporte para interacciones humanas. Desde este punto de vista, en un futuro cercano, la gente estará rodeada por interfaces inteligentes e intuitivas embebidas en los objetos cotidianos que nos rodean y un entorno que reconocerá y responderá a la presencia de individuos de manera transparente. Los ambientes inteligentes están contruidos a partir de sistemas informáticos omnipresentes e interfaces de usuario inteligentes

que permiten interacción social y comunicación con el entorno de manera natural y personalizada.

Las investigaciones actuales sobre entornos inteligentes tratan con interacción social, natural y multimodal [15, 16]. La mayoría de las investigaciones acerca de interfaces sociales están enfocadas al diseño de los llamados agentes corpóreos conversacionales (ECAs, Embodied Conversational Agents) [5]. Estos agentes se hacen visibles en el interfaz como actores animados que representan seres humanos. Algunas veces, consisten simplemente en caras animadas que hablan, mostrando en ocasiones expresiones faciales y, cuando se usa síntesis de voz, aportando sincronización labial. Otras veces tienen una representación gráfica en 3D, con movimientos corporales y faciales complejos. Estos humanoides animados hacen uso de la voz, el gesto, las expresiones, la entonación y otras modalidades no verbales con el objeto de emular la experiencia de la conversación humana cara a cara. La utilidad de estas interfaces de tipo "humanoide" y sus ventajas frente a otros tipos de interfaz sigue discutiéndose [7], pero parece probado [14] que cuando un humano se comunica con un agente de este tipo revela más información de tipo personal, usa un lenguaje más cuidadoso y acepta mejor las sugerencias y recomendaciones. Es por ello que el uso de estos nuevos agentes de interfaz se está extendiendo de forma rápida y hoy en día están siendo utilizados como presentadores virtuales [2,17,20], como guías en visitas a museos o entornos virtuales [20, 23], en aplicaciones educativas [6,12], como vendedores o agentes inmobiliarios [4] o entrenadores personales [3].

El objetivo de este trabajo es mostrar el uso de un agente animado 3D interactivo como nueva

interfaz para el acceso y control remoto de una habitación inteligente real.

Dentro de los entornos inteligentes, el ámbito del hogar y de las aplicaciones domóticas se está configurando como un campo en el que el uso de agentes de interfaz animados puede ser muy efectivo [18]. Ruyter [21] en su trabajo demuestra que si se dota de inteligencia social a la interacción con un entorno de inteligencia ambiental, se crea una percepción positiva de la tecnología del entorno domótico y mejora la aceptación del usuario del empleo de estos entornos.

Sin embargo, el uso de ECAs en este ámbito no está tan generalizado. Destacan dos trabajos, directamente relacionados con el que se presenta aquí. Se trata del "Maior-Domo" de Fagor [9] y del habitante virtual de Kruppa et al [11].

En el caso del habitante virtual, está pensado para ayudar al usuario en un entorno instrumentado complejo. El motor de animación, al contrario del nuestro, es muy simple y está basado en animaciones Flash. Su particularidad e interés radica en la posibilidad de proyectar la salida visual de forma que se pueda "colocar" al ayudante virtual en distintos puntos de la habitación. Por el momento el trabajo se ha centrado en resolver los problemas técnicos de dicha proyección más que en el comportamiento del habitante o su interacción con el usuario, aspecto central del trabajo que aquí se presenta.

En cuanto al "Maior-Domo", es la representación virtual de un controlador domótico pensado para ayudar al usuario a llevar a cabo tareas domésticas como confeccionar la lista de la compra, programar la lavadora o preparar una comida. La idea que subyace es la misma que la de este trabajo: lograr una interfaz natural vía voz que facilite y haga accesible el uso del entorno domótico. La principal diferencia radica en que nuestro trabajo se centra en el control remoto de la instalación domótica y en cuidar los aspectos relacionados con la expresividad en los gestos faciales y la voz, siendo capaz de tener en cuenta aspectos emocionales como se explicará más adelante.

La estructura del artículo es la siguiente: en el apartado 2 se describe el motor que hace posible el uso del agente de interfaz 3D interactivo, mientras que en el apartado 3 se presenta el entorno domótico real puesto en marcha. El apartado 4 presenta el nuevo interfaz propuesto

para el control remoto del entorno domótico, y en el apartado 5 se muestra la aplicación desarrollada. Se finaliza con unas conclusiones y líneas de trabajo futuras.

2. El motor Maxine

Maxine [22] es un motor gráfico para la creación y gestión de entornos virtuales tridimensionales. Dichos entornos pueden incluir actores sintéticos con la capacidad, entre otras, de síntesis de voz y de animación labial sincronizada, lo que los hace especialmente útiles para su uso como interfaces naturales.

Programado íntegramente en C++, está construido sobre un amplio conjunto de librerías de código abierto (o libre). La mayor parte de la aplicación es independiente de la plataforma, sin embargo, la síntesis de voz (basada en tecnología propietaria de Microsoft) hace que la versión actual de Maxine funcione en entorno Windows.

El control de la aplicación se lleva a cabo mediante *scripting*. Pueden ejecutarse comandos desde ficheros, *scripts* cargados al inicio de la aplicación o durante la misma, o bien ser introducidos a través de una consola de texto en pantalla.

2.1 Características generales de Maxine

Maxine es un motor para la creación, organización, manipulación y animación de diferentes elementos que constituyen un escenario virtual, a través de un grafo de escena. Los siguientes elementos pueden ser representados en un grafo de escena:

- Imágenes y textos, que pueden ser mostrados y posicionados como rectángulos orientados en el espacio. Se permiten los principales formatos gráficos (bmp, gif, jpeg, pic, png, rgb, tga, tiff) y se soporta el canal alpha.
- Primitivas geométricas simples pero también modelos geométricos complejos. Maxine importa la mayoría de los formatos populares (3DS, flt, lwo, md2, obj, osg), permitiendo crear escenarios virtuales complejos con alto nivel de detalle.
- Luces simples.
- Sonido de ambiente y 3D.

- Personajes animados. Usan el formato de la librería de animación Cal3D, pero pueden ser generados mediante herramientas de modelado y animación 3D comerciales. Se proporcionan diferentes tipos de animaciones, incluyendo animación secundaria para incrementar el realismo y la expresividad de los actores.
- Actores animados. Son una especialización de los anteriores y están provistos de síntesis de voz y animación facial con sincronización labial, siguiendo el estándar VHML y la especificación MPEG4 para la animación facial.
- Voces sintéticas. Las voces se pueden asociar a un elemento. Se pueden configurar y controlar ciertas características de audio como la selección de una voz concreta, el control de volumen, la velocidad del habla, inserción de pausas, el tono, la enfatización de las palabras, especificación de la pronunciación, etc.

2.2 Interacción en Maxine

La interacción del usuario con Maxine se llevaba inicialmente a cabo vía teclado, mediante a introducción de comandos en el lenguaje de scripting y/o asociando la ejecución de un comando a la pulsación de determinadas teclas.

Con el objeto de conseguir una interacción multimodal con el usuario, posteriormente se ha añadido un sistema de diálogos para la interacción vía voz. Maxine es capaz de reconocer el habla del usuario para llevar a cabo la interacción mediante voz, que permita el diálogo con actores virtuales. Incluye un reconocedor del habla que permite procesar las frases de un usuario y generar respuestas apropiadas en tiempo real, con animación labial sincronizada y gestión del estado emocional de los actores sintéticos (ver Figura 1). El reconocimiento de voz en lengua castellana está soportado gracias a Loquendo ASR (Audio Speech Recognition) [13], un conjunto de librerías comerciales que requieren de la correspondiente licencia. La síntesis de voz en castellano también se ve condicionada por el mismo hecho.

Adicionalmente, Maxine permite crear en tiempo real pequeñas bases de datos con sus correspondientes sistemas gestores que permiten almacenar las respuestas a diferentes preguntas provenientes de los usuarios o del propio sistema Maxine. El funcionamiento de este motor de

búsqueda de respuestas, así como la sintaxis utilizada para definir una base de datos asociada se basa en la metodología de trabajo de AIML. AIML [1] es un derivado del XML cuyo objetivo es definir conocimiento basado en estímulo-respuesta de forma estructurada y lista para ser servido, recibido y procesado en la Web. El intérprete AIML ha sido modificado para incluir comandos o llamadas a ficheros script dentro de las categorías AIML. Esta funcionalidad es la que hace posible la interacción remota con la habitación inteligente que se explica en el apartado 4, donde se detalla el sistema de diálogos.

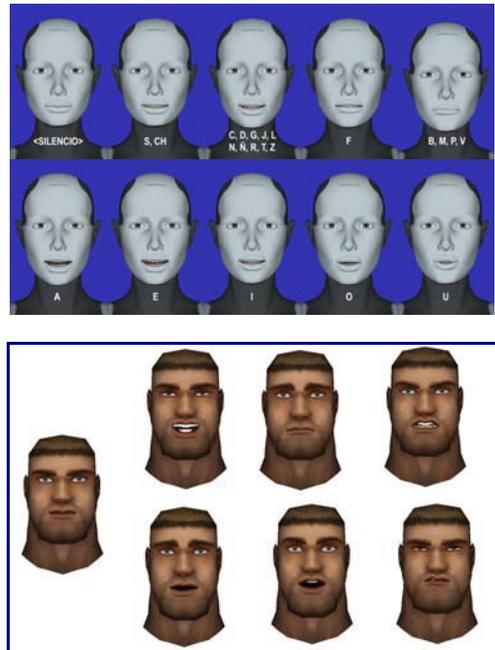


Figura 1. Visemas (representación 3D de fonemas) usados para la sincronización labial en español (arriba). Expresión facial neutral y las seis emociones básicas: alegría, tristeza, enfado, miedo, sorpresa, asco (abajo)

3. Entorno inteligente domótico

3.1. Descripción

Con el objetivo de trasladar al mundo físico las ideas sobre inteligencia ambiental se ha desarrollado en el Departamento de Ingeniería Informática de la Universidad Autónoma de Madrid un entorno inteligente real (ver Figura 2).

Las principales características de este entorno son:

- El control de los dispositivos físicos del entorno se realiza mediante el bus domótico EIB (European Installation Bus), actualmente denominado KNX.
- Para el flujo de información multimedia se emplea una red Ethernet.
- La definición del entorno y los elementos que forman parte de él se realiza de forma estándar y sencilla, basándose en el lenguaje XML.



Figura 2. Entorno inteligente en la Universidad Autónoma de Madrid

- A partir de esta definición se crea un repositorio común, que se denomina pizarra, donde queda representado el entorno, sus elementos, relaciones, estado, flujos de información, aplicaciones, etc.
- La información física de los dispositivos del entorno se encuentra encapsulada en esta capa de pizarra, de tal modo que se puedan cambiar unos dispositivos por otros de igual funcionalidad sin variar las capas superiores.

- Este repositorio almacena información de contexto que puede ser utilizada por el resto de aplicaciones.
- Esta información se encuentra continuamente actualizada de acuerdo con la situación actual del entorno.
- Para modificar el estado de alguno de los elementos del entorno bastará con enviar mensajes estándar de alto nivel a la pizarra. A su vez, para obtener el estado de algún elemento, basta con realizar una solicitud a la misma.

Los elementos físicos que se pueden controlar en un entorno inteligente se pueden clasificar en dispositivos de control y dispositivos multimedia. Entre los dispositivos de control se encuentran los sensores (de presencia, temperatura, etc.) y los actuadores (en luces, puertas, electrodomésticos, etc.). Los dispositivos multimedia engloban a radios, televisores, pantallas de información, micrófonos, altavoces, etc.

Los dispositivos de control se gestionan mediante el bus domótico. Este bus surge como resultado del proceso de estandarización en el control de los elementos de una red domótica. Los dispositivos presentan una arquitectura distribuida. Cada elemento lleva incorporado de forma implícita la información de procesamiento que necesita en microcontroladores, de modo que cada uno funciona de forma independiente a los demás y es el bus el que se ocupa de comunicarlos.

Los dispositivos multimedia están conectados a una red Ethernet. Estas redes resultan estándar, conocidas y ampliamente utilizadas dentro del campo de las telecomunicaciones, por lo que su utilización simplifica el empleo y desarrollo de aplicaciones multimedia dentro del entorno. Ambas redes se armonizan mediante una capa SMNP (Simple Management Network Protocol).

Algunos de los elementos que componen el entorno son:

- Un televisor y una cafetera, con características de control básicas como encendido y apagado.
- Dos altavoces de alta fidelidad que permiten reproducir sonido proveniente de cualquier fuente del entorno, por ejemplo de la síntesis de voz.

- Unas pantallas planas que muestran información personalizada dependiendo de los ocupantes del entorno.
- Un micrófono que se utiliza para registrar la señal de voz.
- Una cerradura electrónica que permite controlar el acceso a la puerta principal del entorno.
- Un detector electromagnético que permite conocer si la puerta se encuentra abierta o cerrada.
- Un conjunto de relés que se emplea para controlar cinco luces del entorno. La primera de las luces corresponde a los fluorescentes situados en el techo. Las otras cuatro se dividen en dos lámparas de pie. Cada una de las lámparas posee un foco y una luz superior regulable en intensidad.
- Interruptores EIB que permiten el control manual de las luces mencionadas en el punto anterior.

3.2. Capa de contexto

La capa de contexto se basa en una estructura global de información que se gestiona utilizando la metáfora de pizarra. A diferencia de otras arquitecturas de tipo pizarra donde la información se almacena en forma de tuplas, la capa de contexto se basa en un grafo dirigido donde los nodos consisten en entidades que representan el contexto del entorno y los vértices representan relaciones entre las entidades. Tal como se describe en [10], esta variante facilita la navegación por parte de los usuarios y aplicaciones -como es el caso de Max- a través de la información contextual.

La flexibilidad y la simplicidad de las arquitecturas de pizarra las hacen una solución muy adecuada para entornos donde la configuración varía frecuentemente.

La capa de contexto se despliega sobre la infraestructura de dispositivos y redes de comunicación que conforman el entorno experimental.

Ésta presenta a los módulos de la capa de aplicación una interfaz común que abstrae la diversidad de dispositivos del entorno físico, de tal forma que cualquier tipo de red de dispositivos que implemente una pasarela a nuestra capa de contexto puede integrarse dentro del sistema.

Los dispositivos físicos se comunican con la capa de contexto, utilizando mensajes estándar en XML y el protocolo HTTP (HyperText Transfer Protocol). Por lo tanto, basta con que la entidad (ya sea un dispositivo físico o una aplicación) que vaya a formar parte del entorno sea capaz de recibir estos mensajes utilizando el protocolo HTTP.

Las entidades disponen de la API de comunicación proporcionada por el sistema. Esta API permite la interacción de los dispositivos con el middleware. A su vez, nuevos dispositivos pueden crear nuevas APIs para lenguajes específicos. En todo caso, el protocolo y los mensajes son estándar y abiertos, de modo que resulta sencillo desarrollar nuevas APIs de comunicación.

4. Max: el nuevo interfaz natural

El objetivo de este trabajo ha sido, aprovechando las características de Maxine, crear un agente (Max) que haga de nuevo interfaz interactivo para el acceso y control remoto de la habitación inteligente. El interfaz debe permitir dialogar de forma natural con el agente (Max), al cual se le debe poder pedir que realice operaciones o consultas sobre los distintos dispositivos de la habitación inteligente.

4.1. Funcionalidades

Los dispositivos sobre los que tiene sentido actuar de forma remota en la habitación prototipo son:

- Luces normales
- Luces regulables (nuestro interfaz permite subir o bajar la intensidad entre cero y la intensidad máxima)
- Cafetera
- Televisión

Todos estos dispositivos pueden encenderse, apagarse o consultar si están encendidos o apagados.

Además también se puede actuar de forma remota sobre:

- Sensor de puerta (la puerta no puede abrirse ni cerrarse remotamente, solo consultar)

VIII Congreso de Interacción Persona-Ordenador

- Personas y Habitación: en todo momento se puede consultar si hay alguien en la habitación y de quién se trata.

4.2. Integración/adaptación de Maxine

Para proporcionar el interfaz interactivo que permita dialogar con el avatar se ha utilizado Maxine y su sistema de preguntas y respuestas. El acceso a la pizarra de forma remota se hace mediante urls, ya que resulta fácil utilizar direcciones web en scripts que, como se ha comentado, pueden ejecutarse dentro del sistema de preguntas y respuestas de Maxine.

Por tanto el trabajo realizado ha consistido básicamente en la creación de:

1. Un script para ejecutar en Maxine que cargue un actor virtual y un plano de la habitación, active y gestione el sistema de preguntas y respuestas y el de reconocimiento de voz, etc.
2. Una gramática para el reconocedor de voz que contenga todas las palabras que este debe entender (tele, puerta, abrir, encender, etc.)
3. Un fuente AIML para la creación de la base de datos con las distintas respuestas. Las respuestas pueden ser respuestas verbales simples o respuestas

que impliquen utilizar un script para la comunicación con el sistema de domótica mediante url, ya sea para consultar o para modificar alguna propiedad.

4. Distintos archivos fuente que contengan todos los scripts necesarios para realizar las operaciones y consultas permitidas. Estos scripts serán invocados desde AIML.

4.3 Comunicación vía voz con el agente

En la comunicación vía voz con el agente (Max) se distinguen tres etapas (ver Figura3):

1. El usuario formula una orden ó pregunta. El sonido o audio generado por el usuario es recogido por el micrófono y la tarjeta de sonido. El reconocedor de voz, ayudado con una gramática que le indica el conjunto de palabras que puede reconocer (en este caso palabras relacionadas con la domótica, por ejemplo puerta, apaga, enciende, etc.), convierte el audio de entrada en una cadena de texto con las palabras pronunciadas por el usuario.

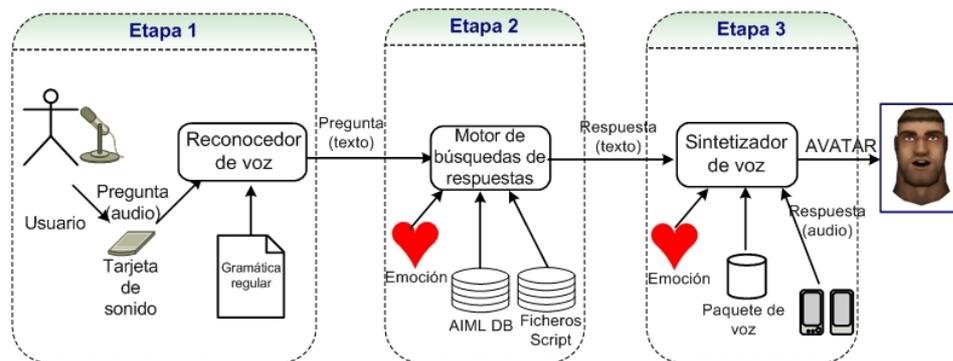


Figura 3. Etapas en el proceso de comunicación con el avatar

2. El motor de búsqueda de respuestas procesa el significado de la entrada y le asigna una respuesta a esa pregunta, buscándola en la base de datos AIML o generándola a partir de la ejecución de un script (desde estos script por ejemplo es desde donde se utilizan las url correspondientes para consultar y modificar propiedades de los dispositivos de la habitación). Esta respuesta todavía está en modo texto.
3. El sintetizador de voz, utilizando un paquete de voz en castellano, extrae de la cadena de texto los visemas correspondientes a cada palabra, necesarios para realizar la sincronización labial del avatar con la síntesis de la cadena de texto. De esta forma el usuario puede oír la respuesta.

modificando la modulación de la voz trabajando sobre los parámetros de tono, escala de la frecuencia, volumen y velocidad del habla. El agente tiene un estado emocional, gestionado por una variable. El sistema permite tener en cuenta dicho estado no sólo en la síntesis de voz sino también en la generación de las respuestas, permitiendo establecer diferentes respuestas según dicho estado. Se trata de una funcionalidad de Maxine que en esta aplicación no ha sido todavía suficientemente explotada.

5. Aplicación

Al iniciar la aplicación el usuario puede ver al agente Max a la izquierda y a la derecha un plano de la habitación donde se puede ver la disposición de las distintas luces, la tele, la cafetera, etc. (ver Figura 4).

Para evitar que la voz suene artificial, ésta ha sido dotada con una componente emocional. La generación de emociones en tiempo real se realiza



Figura 4. Nuevo interfaz desarrollado

VIII Congreso de Interacción Persona-Ordenador

A continuación, Max pronuncia el siguiente texto:

*“Hola, me llamo Max.
Si quieres saber todo lo que puedo hacer solo tienes que preguntármelo.
Y si ya lo sabes puedes decirme lo que quieres que haga.”*

A partir de ahí se puede dialogar con él (ver Figura 5), por ejemplo se le puede preguntar que puede hacer:

Usuario: *“¿Qué puedes hacer?”*
Max: *“Puedo decirte como están las luces, la cafetera o la tele, o si la puerta esta abierta o cerrada. Puedo apagar luces, encender luces, y subir y bajar las luces regulables, también puedo encender y apagar la tele y la cafetera, o decirte quién hay en la habitación. ¿Qué quieres hacer?”*

Se le puede dar órdenes como por ejemplo:

Usuario: *“Enciende la luz uno”*
Max: *“La luz ya esta encendida”*

Usuario: *“Enciende la luz uno”*
Max: *“Hecho”* (la luz uno pasa a estar encendida)

Usuario: *“Enciende la luz uno”*
Max: *“Debido a un error no he podido encender la luz”*
Usuario: *“Abre la puerta”*
Max: *“La puerta no puede abrirse ni cerrarse desde aquí”*

Usuario: *“Sube la luz tres”*
Max: *“Listo”* (Aumenta la intensidad de la luz tres)

El interfaz permite ver los cambios que producen estas órdenes a través de la cámara web.

También se puede consultar el estado de los distintos dispositivos y las personas que están en la habitación:

Usuario: *“¿Cómo esta la puerta?”*
Max: *“Está cerrada”*

Usuario: *“¿Está la tele encendida?”*
Max: *“Esta apagada”*

Usuario: *“¿Hay alguien?”*
Max: *“Hay dos personas”*

Usuario: *“¿Quiénes son?”*
Max: *“Manuel, Germán”*



Figura 5. Usuaría conversando con el agente Max.

6. Conclusiones y Trabajo futuro

Se ha desarrollado un nuevo interfaz para el control remoto vía web de una sala domotizada. El nuevo interfaz se basa en el uso de un agente virtual 3D con el que se interacciona vía voz en lenguaje natural. A través de dicho agente de interfaz el usuario no sólo obtiene información sobre el estado de los diferentes elementos (puerta, luces, televisor,...) sino que puede cambiarlo. Adicionalmente, a través de una webcam puede visualizar los cambios que realiza en la estancia. El usuario no debe memorizar ningún comando específico ni su uso requiere ningún tipo de entrenamiento. Frente al interfaz original, basado en un formulario, la interacción es ahora mas natural y, por tanto aumenta la accesibilidad del sistema, ampliando su uso a usuarios con discapacidad o de edad avanzada. Una cuestión a considerar sería, siguiendo el ejemplo del sistema inteligencia

ambiental Embassi [8], el desarrollo de un módulo de presentación que decida qué modalidad se va a utilizar como salida y se asegure de que ésta sea coherente y consecuente.

El agente 3D ha sido desarrollado gracias a un sistema, Maxine, orientado a la generación y control de agentes 3D dotados de interacción multimodal. Se ha hecho especial hincapié en los aspectos expresivos y emocionales tanto de la animación como de la voz.

En cuanto al trabajo futuro, además de enriquecer las funcionalidades del sistema, se está estudiando la adaptación del interfaz para el acceso y control por medio de otros dispositivos como teléfonos móviles o PDAs.

Queda así mismo pendiente una adecuada evaluación del interfaz desarrollado. Los agentes Maxine han sido evaluados en su uso como interfaz en otras aplicaciones de tipo educativo desarrolladas; los resultados han sido buenos en cuanto a su naturalidad corporal y facial y algo peores en lo referente a la voz sintética del agente. En cualquier caso, ha sido positiva en cuanto a su recomendación de uso en aplicaciones de educación a distancia. Sin embargo, una evaluación de su uso en el entorno domótico es necesaria.

Agradecimientos

Los autores quieren agradecer a David Anaya y Sergio García-Massip su trabajo en el motor Maxine.

Este trabajo ha sido parcialmente financiado por el Gobierno de Aragón gracias al convenio WALQA (ref. 2004/04/86) y al proyecto N° CTPP02/2006 así como por la Dirección General de Investigación a través de los proyectos TIN2004-07926 y TIN2004-03140.

Referencias

- [1] Artificial Intelligence Markup Language (AIML) v1.0.1 A.L.I.C.E. AI Foundation Working Draft. Autor: Noel Bush. 25 October 2001 (rev 006). Publicado en: <http://www.alicebot.org/TR/2001/WD-aiml/>
- [2] Baus, J., Butz, A., Krüger A. Incorporating a Virtual Presenter in a Resource Adaptive Navigational Help System. Workshop notes of the Workshop on Guiding Users through Interactive Experiences, Paderborn, Germany, April, 2000
- [3] Bickmore, T. Relational Agents: Effecting Change through Human-Computer Relationships. PhD Thesis, Media Arts & Sciences, MIT, 2003
- [4] Cassell, J., Bickmore, L., Campbell, Vilhjalmsón, H., Yan H. Human Conversation as a System Framework: Designing Embodied Conversational Agents. In Embodied Conversation Agents, J. Cassell, J. Sullivan, S. Prevost, E. Churchill (eds), pp. 29-63, 2000
- [5] Casell, J., Sullivan, J., Prevost, S., Churchill E (eds). Embodied Conversational Agents, Cambridge: MIT Press, 2000
- [6] Cerezo, E., Baldassarri, S., Serón, F. The use of interactive animated agents for teaching. Proc. ICIE 2007. 3rd International Conference on Interdisciplinarity in Education, 2007.
- [7] Creed, C., Beale, R. Embodied Interfaces: The Next Generation of HCI? Proc. of the Workshop on The Next Generation of HCI in cooperation with the Conference on Human Factors in Computing Systems (CHI2006), 2006
- [8] Elting, C., Rapp, S., Möhler, G., Strube, M. Architecture and Implementation of Multimodal Plug and Play. Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI-PUI '03), Vancouver, Canada, 2003
- [9] Gárate, A., Herrasti, N., López, A. GENIO: An Ambient Intelligence application in home automation and entertainment environment. Proc. Joint sOCEUSAI Conference, Grenoble, 2005
- [10] Haya, P. A., Montoro, G., Xavier Alamán. A prototype of a context-based architecture for intelligent home environments. Proc. of the International Conference on Cooperative Information Systems (CoopIS 2004), Larnaca, Cyprus. Lecture Notes in Computer Science (LNCS), volume number 3290, pp. 477-49, 2004
- [11] Kruppa, M., Spassova L., Schmitz M. The virtual room inhabitant -Intuitive interaction with intelligent environments. Artificial Intelligence 2005, Lecture Notes in

VIII Congreso de Interacción Persona-Ordenador

- Artificial Intelligence 3809, pp. 225-234, Springer Verlag, 2005
- [12] Lewis, W., Rickel, J. W. and Lester, J. C. Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, vol. 11, pp. 47-78, 2000.
- [13] Loquendo
<http://www.loquendo.com/es/index.htm>
- [14] Maldonado, H., Hayes-Roth, B. Toward cross-cultural believability in character design. In *Agent Culture: Human-agent interaction in a multi-cultural world*. S Payr and R. Trapp (eds), pp. 177-196, 2001
- [15] Nijholt, A. Multimodal interactions in ambient intelligence, in W. Verhaegh, J. Korst, E. Aarts (eds). *Algorithms in ambient intelligence*, Chapter 2. Phillips Research Book Series. Dordrecht: Kluwer Academic Publishers, 2003.
- [16] Nijholt, A. Where computers disappear, virtual humans appear, *Computer & Graphics* n° 28, pp. 467-476, 2004.
- [17] Nijholt, A., van Welbergen, A., Zwier, J. Introducing an embodied virtual presenter agent in a virtual meeting room. *Proc. of the 23rd IASTED International Multi-Conference Artificial Intelligence and Applications*, Innsbruck, Austria, pp. 579-584, 2005
- [18] Nijholt, A. Google Home: Experience, Support and Re-Experience of Social Home Activities. *Information Science, Special Issue on Ambient Intelligence*, Elsevier, 2007 (en prensa)
- [19] Rist, T., André, E., Baldes, S., Gebhard, P., Klesen, M., Kip, M., Rist, P., Schmitt, M. A Review of the Development of Embodied Presentation Agents and Their Application Fields in Life-Like Characters: Tools, Affective Functions, and Applications, in Helmut Prendinger and Mitsuru Ishizuka (eds), *Springer Series on Cognitive Technologies*, pp. 377-404, 2003
- [20] Rocchi, C., Stock, O., Zancanaro, M., Kruppa, M., Krüger A. The Museum Visit: Generating Seamless Personalized Presentations on Multiple Devices. *Proceedings of the 9th International Conference on Intelligent User Interfaces* IUI'04, January 13-16, Madeira, Portugal (2004)
- [21] Ruyter, B., Saini, P., Markopoulos, P. van Breemen, A. Assessing the effects of building social intelligence in a robotic interface for the home. *Interacting with computers*, 17 (5) , pp. 522-541, 2005
- [22] Seron, F., Baldassarr, S., Cerezo, E. MaxinePPT: Using 3D Virtual Characters for Natural Interaction. *E. Proc. 2nd International Workshop on Ubiquitous Computing & Ambient Intelligence*, pp. 241-250, 2006.
- [23] Yuan, X., Chee, Y.S. Design and evaluation of Elva: an embodied tour guide in an interactive virtual art gallery. *Computer Animation and Virtual Worlds* 16, pp.109-119, 2005