Validación de una estrategia de interacción de un agente corpóreo conversacional a través de la técnica del mago de Oz

Daniel Martínez García[1]
(danielmtz601@gmail.com)

Paul Craig[1]
(p.craig@ mixteco.utm.mx)

Néna Roa Seïler[1,2]
(n.roa-seiler@napier.ac.uk)

Ariadna Benítez Saucedo[1]
(aribesa@gmail.com)

[1]Universidad Tecnológica de la Mixteca
División de Postgrado
Huajuapan de León, Oaxaca, C.P. 69000

[2]Edinburgh Napier University
42 Colinton Road
Edinburgh, United Kingdom, EH10 5BT

RESUMEN
Los agentes corpóreos conversacionales son interfaces prometedoras en la interacción entre los humanos y los sistemas computacionales. Sin embargo para ser aceptables como los humanos virtuales que pretenden ser, necesitan ser capaces de mostrar una de las características que definen a los seres humanos: entender y expresar emociones durante una interacción. Este artículo presenta la implementación y la creación de un sistema de diálogo hablado humano-computadora aplicando la técnica del "Mago de Oz" (WoZ), la cual nos permitiría la obtención de un corpus para validar y mejorar el sistema. Nuestra plataforma multimodal está compuesta por un agente con características antropomórficas: caracterización y vestimenta de una mujer adulta en 3D, que posee expresiones gestuales emocionales, voz sintética, y combinación de síntesis de voz con expresiones faciales. Las características anteriormente mencionadas permitirán a los usuarios que interactúen con el sistema, haciéndoles creer que se encuentran conversando con un agente corpóreo inteligente. Para evaluar nuestro sistema de diálogo se necesita un número importante de experimentos de WoZ con el fin de compararlos y en los que se involucren diferentes comportamientos en el avatar por ejemplo: compasivo, alentador, inquisitivo, en espera, en escucha, interrumpido, confundido, sorprendido. En el presente artículo describimos el desarrollo del sistema, su gestación, su estructura, la metodología a seguir para la obtención del corpus y proponemos una interfaz gráfica.

Palabras clave del autor
Mago de Oz; agentes corpóreos; conversacionales; expresiones gestuales; sistema de diálogo hablado.

MexIHC 2012, October 3-5, 2012, Mexico City, Mexico.

Clasificación de ACM
H.1.2.a Human factors

INTRODUCCIÓN
Los agentes conversacionales que interactúan con los humanos a través del diálogo hablado, se han convertido probablemente en la comunicación más cercana a la utilizada por los humanos de manera intuitiva. Esta posee la voz, el tono de ésta, las expresiones gestuales, expresiones corporales, la mirada, el turno de la conversación. Todas estas características han formado parte del interés por parte de la ciencia de la Lingüística Computacional desde sus inicios.

En la creación de nuestro proyecto es necesario extraer una cantidad considerable de corpus. Para obtenerlo hemos elegido la técnica del WoZ, estas pruebas serán aplicadas en el laboratorio de usabilidad dentro de las instalaciones que se encuentran en la Universidad Tecnológica de la Mixteca, esto nos permitirá mejorar el diseño de la interfaz de administración, evaluar el sistema del diálogo, con la finalidad de obtener una medición sobre la expectativa creada en los usuarios por parte del agente y del sistema en general.

La técnica de WoZ ha sido aplicada desde principios de la década de los ochenta, esta permite la recolección de corpus de diálogos persona-computador. El computador es simulado por una persona oculta que realiza todas o algunas de las funciones que realizará el sistema de diálogo definitivo [6]. De esta manera se desea obtener resultados que nos puedan ayudar a mejorar tanto el comportamiento del agente como la interacción humano-computadora de manera satisfactoria, así como agilizar la respuesta del sistema durante la fase de interacción.

IMPORTANCIA DE LOS AGENTES CONVERSACIONALES CORPÓREOS (ECA)
En la actualidad un buen número de investigaciones sobre futuras interfaces corresponden a los Agentes Animados Conversacionales o Agentes Corpóreos Conversacionales

(en Inglés Embodied Conversational Agents), que por comodidad en este documento llamaremos ECAs.

Los ECAs ofrecen la posibilidad de combinar varios elementos utilizados en la comunicación cara a cara como los diálogos y las expresiones gestuales. Han sido ya empleados para mejorar la interacción, gracias a ellos están emergiendo nuevas teorías de diseño de interfaces más naturales y confortables aunque queda aún un largo camino para mejorar la interacción gracias a la incorporación de elementos de la comunicación no verbal [9].

Hay algunas situaciones de interacción que un ECA podría ayudar y tener un efecto positivo. Puede por ejemplo mejorar el manejo oportuno y eficiente del turno. El lenguaje corporal y la expresividad de los agentes es importante no únicamente para reforzar el mensaje hablado, si no también cómo nos menciona Cassell [4], para regular el flujo del diálogo, así en caso de producirse un error, la utilización de un ECA, puede ayudar a retomar el curso del diálogo y así recuperarse del error, usando las características antropomórficas y expresiones faciales como corporales, a permitiendo un diálogo fluido. Los ECAs también pueden mejorar en la recuperación de errores. El proceso de reconocimiento permitiendo recuperación de errores que usualmente conducen a un seguro desagrado por parte del usuario frustrado [11]. También puede ayudar a reducir la frustración y debe hacerlo de la manera más efectiva que pueda [8]. Al igual que al entendimiento del estado del diálogo, un problema que se presenta continuamente, es cuando el usuario no sabe si el sistema está trabajando normalmente [12]. Los ejemplos mencionados anteriormente muestran como un ECA puede evitar y reducir problemas con el sistema.

TRABAJOS RELACIONADOS

A continuación mostramos trabajos similares al nuestro, en los cuales hacen uso de agentes. AVARI (Animated Virtual Agent Retrieving Information) [5] es una recepcionista virtual que brinda información usando un lenguaje natural y expresiones faciales, tono e ingeniosidad verbal, resultando accesible y efectivo; pertenece a la facultad de ciencias de la computación de la Universidad del Norte de Carolina en Charlotte (UNCC). FREUBOT [7] es un ECA que interactúa con estudiantes sobre conceptos de Sigmund Freud, adquiridos de la biografía del mismo personaje. Este ECA no fue programado para analizar o ayudar a los usuarios, pero si para platicar sobre las teorías de Freud. NEVA, ayuda a los lectores a buscar sus libros en la biblioteca. Es un avatar programado para asignar una personalidad a los usuarios, conociendo sus intereses y comportamientos. Ella fué desarrollada para el Centro de Documentación McLuhan de ISNM (*International School New Media*) en la Universidad de Lübeck, Alemania [3]. TQ-BOT es un asistente virtual que fue creado para ayudar a estudiantes y tutores asignándoles un proceso de aprendizaje con diferentes formas con preguntas y tareas usando las plataformas de aprendizaje Moodle o Claroline [10]. SAMIR [1] es un agente en 3D capaz de ayudar a los usuarios cuando buscan una tienda de libros en línea. Es capaz de aprender reglas de comportamiento para mejorar incluso el rendimiento de sus proposiciones. SUSANNA [2] es una ECA que habla con los usuarios desde una tienda de libros en línea (BOL). SUSANNA colecta las necesidades, anhelos e intereses, de los usuarios estableciendo su perfil para hacerles recomendaciones de libros.

METODOLOGÍA DEL MAGO DE OZ A UTILIZAR

Muchos estudios han mostrado que los usuarios reaccionan de manera diferente, cuando interactúan con agentes sociales. En este proceso, influye la personalidad propia de cada persona, así como el ambiente en el cual las personas se encuentran inmersas.

Para obtener el corpus de diálogo, lo primero que se tiene pensado realizar es un *focus group,* en dónde debemos explicar el estudio que nos encontramos realizando a los posibles usuarios (profesores de idiomas que se encuentran en labores dentro de la Universidad Tecnológica de la Mixteca), y por consiguiente se les brindarán un cuestionario para recolectar sus pasatiempos, así como palabras o expresiones que ocupan de manera más cotidiana dependiendo de la situación o estado de ánimo en la que se encuentren, y después de eso poder agendar las fechas para la aplicación de la prueba del WoZ.

El segundo paso consistirá en aplicación de la prueba y estará formada por las siguientes etapas. Durante la etapa del *breve explicación***,** se informará al usuario del propósito de las pruebas y las tareas que deben llevar acabo (se montarán escenarios uno sobre sus pasatiempos y el otro sobre su estado emocional en ese día). El *fase del diálogo e interacción*, será el momento en que los participantes mantienen un diálogo con el sistema. El *cuestionamiento final*, es la etapa final en dónde se obtiene la impresión global del sistema, así como sugerencias. A partir de esto podremos analizar el corpus y mostrar el resultado de los experimentos.

Instrumentos a utilizar

Como mencionamos anteriormente, los experimentos del WoZ, se realizarán en un laboratorio de usabilidad, ubicado dentro del campus de la Universidad Tecnológica de la Mixteca. Este laboratorio, cuenta con videocámaras de registro (importante para la adquisición del corpus, a través de la transcripción), además que cuenta con un muro divisional en la cual se puede observar a los participantes del experimento, sin poder ser visto y/o escuchado. Los instrumentos que utilizaremos para esto serán: una computadora portátil, un monitor adicional conectado a la computadora, una diadema con audífonos y micrófono, mesas, sillas confortables y videocámaras con el respectivo equipo a utilizar para la grabación experimentos.

ARQUITECTURA DE NUESTRO MAGO DE OZ

Nuestra tarea consiste en facilitar la manipulación del panel de administración para el “mago”. Para poder hacer uso y

manejo del panel de administración, debe tener los siguientes requisitos: Manejo gramatical y prosodial de la lengua inglesa; conocimientos en la interacción con sistemas de diálogo; conocimiento del desarrollo del sistema y de las partes que integran el panel de control (previamente instruido).

El sistema se basa en un entorno gráfico el cual está conformado por secciones del panel de administración del WoZ que son los botones radiales, el avatar y la caja de texto. Los botones radiales representan cada una de las 8 posibilidades de interacción emocional que puede expresar el ECA: compasivo, alentador, inquisitivo, en espera, en escucha, interrumpido, confundido, sorprendido. Estos estados corresponden a comportamientos que el ECA va mostrar en función de la interacción que se esté realizado, por ejemplo si durante la conversación el ECA es interrumpido deberá mostrar una expresión facial y una expresión vocal que animen al usuario a continuar la conversación. Al pulsar con el ratón de la computadora sobre alguno de ellos, aleatoriamente se ejecuta una emoción. Actualmente el sistema cuenta con atajos de teclado debido a la necesidad de agilizar el procesamiento de respuesta de parte del mago. El avatar se encuentra en la parte derecha. Este es un agente animado en 3D (conocida bajo el nombre de "Samuela"), la cual presenta expresiones emocionales que le indique el "mago", de tal forma que simule a un humano. Este agente animado fue diseñado e implementado con el software Haptek editor. El agente tiene como características antropomórficas: ojos azules, cabello rubio, lacio y largo, saco color blanco. La caja de texto es la parte de la interfaz en donde se escribe el texto que se desea enviar al servidor para convertirla en diálogo hablado (De Texto a Audio). Señalamos además que se encuentran dos botones en la parte derecha de la caja de texto, uno ejecuta la función que hace hablar a nuestro avatar (es decir lo que se encuentra escrito), y el otro ejecuta la expresión oral escrita en la caja de texto más la expresión facial (elegida) del avatar de manera sincronizada.

Estos elementos gráficos del panel de administración del WoZ se puede ver más abajo (la Figura 1). El panel ha sido desarrollado a través de los lenguajes HTML, Javascript y Java Server Pages (JSP). El diseño de la interfaz gráfica del panel de administración fue elaborado usando hojas de estilo en cascada (mejor conocido como CSS).

Para ejecutarse la aplicación hace uso de un servidor Tomcat 5 (Apache). Éste se arranca a través de scripts elaborados en el Shell de Windows, que a su vez ejecuta dos archivos ejecutables .JAR de Java.

El sistema actualmente se encuentra en pruebas en una máquina con un sistema operativo Windows Vista. La aplicación del panel de administración, así como el agente conversacional en 3D sólo puede ejecutarse en el navegador Internet Explorer, debido al plugin de haptek de Microsoft, que es un software privativo.

El agente en 3D, como mencionamos anteriormente está diseñado en Haptek editor, un software que nos permite crear avatares permitiéndonos personalizarlos tanto las expresiones faciales como la posición y la forma de las cejas, la boca, etc. En nuestro caso el agente puede ejecutar 8 estados en los que el ECA puede aparecer enfrente de los usuarios por ejemplo el ECA puede mostrarse comprensivo, alegre, en escucha, neutral/ocioso, en modo de cuestionamiento, confundido, curioso, sorprendido.

Estos estados están compuestos de expresiones faciales y de expresiones vocales en correspondencia con la estrategia de interacción que el ECA debe mostrar. La ejecución de cada emoción sólo se puede realizar por medio de la captura de pulsar el botón izquierdo del ratón sobre cada elemento de los botones radiales (*radio buttons*) que corresponden a cada emoción, o incluso cambiando el foco de un botón radial a otro, a través del teclado. En la última implementación hecha a la aplicación se pueden ejecutar presionando un combinado de teclas (*shortcuts*), que ayudan a mejorar el dinamismo y la rapidez de respuesta en la aplicación.

Con respecto a la voz sintética que ocupa el agente, ésta es elegida y sintetizada a través del software SAPI 4 (Figura 2). El cual transforma texto a audio automáticamente. Este programa es el que configura la voz del avatar, en nuestra aplicación la configuración aplica la voz de "Elizabeth" que es una voz femenina con acento de Inglés Británico. SAPI 4 ofrece ciertas ventajas particulares tales como la opción de sincronización de los movimientos y expresiones faciales con la voz del avatar; la síntesis de la voz del avatar, es lo más cercana a una voz humana, con respecto a otros sintetizadores; acepta diálogos hablados en diferentes idiomas. Mientras que con el software TTS (*Text-to-Speech*), es una de las tecnologías más maduras en lo que se refiere a reconocimiento de diálogo textual a audio.

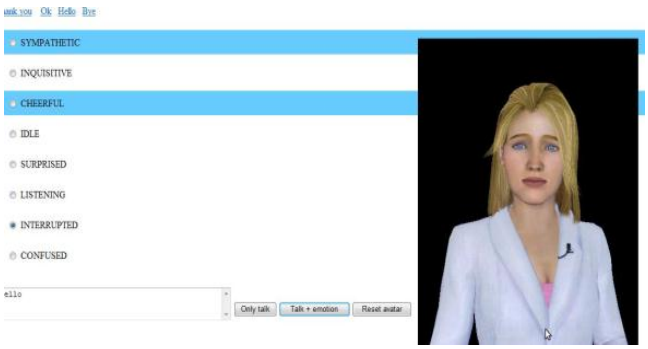

Figura 1. Interfaz del panel de administración.

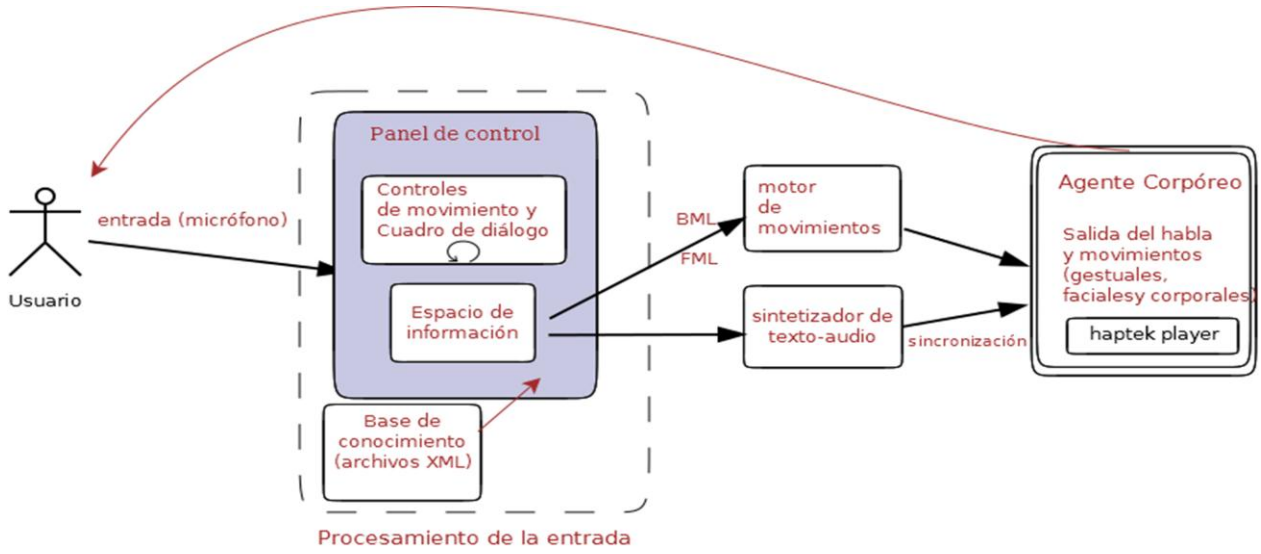

Figura 2: Arquitectura del Mago de Oz.

CONCLUSIÓN

De acuerdo al trabajo emprendido hasta el momento, estamos conscientes de los problemas que se puedan suscitar dentro de la aplicación de cada una de los experimentos del Mago de Oz para con los participantes, como en todo experimento pueden suceder problemas técnicos, así como no se pueden descartar problemas con el sistema (software y hardware), o en su defecto con el manejo del panel de administración. El panel, actualmente se encuentra en una versión estable, sin embargo se mejorará para dar una buena imagen y mayor confortabilidad al administrador que lo use. Para mejorar la interfaz gráfica se aplicará un estilo a través de la codificación de hojas de estilo (CSS), en conjunción con el lenguaje HTML. Y en futuro cercano se tiene pensado crear una base de conocimiento que pueda ayudar al administrador en el manejo del avatar y en situaciones o escenarios determinados.

REFERENCIAS

1. Abbattista, F., Catucci, G., Semeraro, G., Zambetta, F. SAMIR: A Smart 3D Assistant on the Web. *PsychNology Journa*l (2004), 2(1), 43-60.
2. Abbattista F., Lops P., Semeraro G., Andersen V., Andersen H. Evaluating virtual agents for e-commerce. In: *Workshop Embodied conversational agents*. AAMAS (2002), Bologna, Italy.
3. Ahad, A. *Neva: A Conversational Agent Based Interface for Library Information Systems*. Master's thesis, University of Lübeck, Germany, June 2005.
4. Bickmore, T., Cassell, J., Van Kuppevelt, J., Dybkjaer, L. and Bernsen, N. (eds.), *Natural, Intelligent and effective Interaction with Multimodal Dialogue Systems*, chapter Social Dialogue with Embodied Conversational Agents. Kluwer Academic, 2004.
5. Cairco, L., Dale-Marie, W., Fowler, V., LeBlanc, M. AVARI: animated virtual agent retrieving information, in Proc. *The 47th Annual Southeast Regional Conference* (2009), March 19-21, 2009, Clemson, South Carolina.
6. Dahlbäck, N., Jönsson, A. y Ahrenberg, L. Wizard of Oz Studies – Why and How. En: Gray, W. D., Hefley, W. E. y Murray, D. Editores, in Proc. *International Workshop on Intelligent User Interfaces (IUI 1993)*, Orlando, 1993. 193-200.
7. Heller, R. B., Procter, M., Mah, D., Jewell, L., Cheung, B. Freudbot: An Investigation of Chatbot Technology in Distance Education. In Proc. *The World Conference on Multimedia, Hypermedia, and Telecommunications* (2005).
8. Hone, K., Animated Agents to reduce user frustration, in *The 19th British HCI Group Annual Conference*, Edinburgh, UK, 2005.
9. Massaro, D. W., Cohen, M. M., Beskow, J., and Cole, R. A., Developing and evaluating conversational agents. In *Embodied Conversational Agents MIT Press*, Cambridge (2000), MA, 287-318.
10. Mikic, F. A., Burguillo, J. C., Llamas, M. TQ-Bot: An AIML-based Tutor and Evaluator Bot. Journal of Universal Computer Science (JUCS 2010), Verlag der Technischen Universität Graz, Austria, 1486-1495.
11. Oviatt, S. & VanGent R., Error resolution during multimodal humancomputer interaction. In Proc. *International Conference on Spoken Language Processing*, 1 (1996), 204-207.
12. Oviatt, S. Interface techniques for minimizing disfluent input to spoken language systems. In *Proc. CHI'94*, ACM Press (1994), 205-210.