I.S.S.N.: 1138-2783

SINCRONIZACIÓN DE LABIOS: MÉTODO SIN VISEMAS

(LIP-SYNCHRONIZATION: A METHOD WITHOUT THE USE OF VISEMES)

Francisco Cabiedes
Ildikó Pelczer
Fernando Gamboa
Universidad Nacional Autónoma de México (México)
Javier Bretón
Steve Rodríguez
Dirección de Investigación e Innovación-Enciclomedia, ILCE (México)

RESUMEN

En el trabajo presente se describe un método simplificado como solución para el problema de sincronización de labios con archivos audio. El algoritmo recibe como entrada el archivo de audio en formato estandarizado (WAV) y regresa un archivo de texto con valores que indican amplitudes promedios de la información audio correspondiente a una fracción de segundo predefinido. Para lograr la sincronización del audio con movimientos de la boca, los valores tal determinados se correlacionan con archivos de animación. La solución adoptada permite la automatización, por una parte, del análisis de un archivo de sonido registrado por cualquier usuario y, por otra, la asociación del archivo audio con una secuencia de archivos de animación. El método fue empleado con éxito en un proyecto con propósitos educativos de gran escala en México.

Palabras clave: sincronización de labios, animación 2D/3D, análisis de archivos audio.

ABSTRACT

In the present article we describe a simplified method to resolve the problem of lip synchronization with a previously recorded audio file. The algorithm receives as input the audio file, in a standard format (WAV) and returns as result a text file that contains average amplitude values corresponding to a predefined fraction of a second. In order to achieve the synchronization, the values such defined are correlated with lip animation sequences. The adopted solution allows an automatic production process that is, on one hand, the analysis of an arbitrary audio file recorded by a user and, on other hand, the

association of the file with a sequence of animation files. The method was successfully employed in a large scale educational project in Mexico.

Key words: lip synchronization, 2D/3D animation, audio file analysis.

En la última década ha crecido considerablemente el número de sistemas tutoriales inteligentes. Una categoría de estas aplicaciones incluye *Agentes Pedagógicos*: elementos de software tipo Agente (Hewitt, 1977), que tienen como referencia gráfica a un Avatar[1]. Un Avatar suele identificarse como una representación gráfica que se auto-atribuye un usuario en el ciberespacio. Sin embargo, un Avatar puede ser también la representación gráfica de un Agente computacional, que tiene como objetivo instruir al usuario en el uso del software o simplemente apoyarlo en su tarea[2].

Uno de los factores de éxito de esta clase de Avatares es que permiten una relación *cara a cara* entre el usuario y el sistema, o entre varios usuarios, dando origen así a una *personalidad digital.*

Un agente animado cuenta con movimientos, expresividad emocional y, en muchas ocasiones, incluye comunicación verbal. Por supuesto, estas características no están presentes en la misma medida en todos los trabajos reportados en la literatura. Por ejemplo, en el trabajo de Conati (Conati; Zhao, 2000), la información se transmite mediante cuadros de texto, igual que en ADELE (Shaw et al., 1999), mientras que HERMAN (Lester, Stone; Stelling, 1999), COSMO (Lester et al., 1999) y PPP PERSONA (André, Rist; Muller, 1999) son agentes pedagógicos con movimientos complejos, expresividad emocional y comunicación verbal.

Diversos estudios han confirmado el efecto positivo que la presencia de estos acompañantes virtuales tiene sobre el desempeño de un estudiante. En Lester et al. (1997a), los autores presentan los resultados de una evaluación a gran escala del impacto pedagógico de esta clase de agentes, mientras que en Lester et al. (1997b) se estudia el efecto de la presencia de Avatares sobre la motivación del estudiante. Las principales conclusiones de estos trabajos son las siguientes:

- Los estudiantes que utilizaron medios de aprendizaje que incluían un agente pedagógico animado mostraron un avance estadísticamente significativo desde pre-test al post-test.
- En los experimentos con agentes que emplean modalidades tanto visuales (despliegue gráfico) como verbales (habla), los estudiantes mostraron

una mejora considerablemente superior en solución de problemas que en experimentos con agentes sin habla y estáticos.
- El agente pedagógico animado tiene un efecto de fuerte motivación para el estudiante.

Sin embargo, a pesar de la complejidad de los medios (verbal, visual, afectivo) utilizados en los agentes pedagógicos animados para transmitir información, dichos sistemas presentan limitaciones:

- Toda la información que se transmite es predefinida por los creadores de los sistemas al momento de diseñar el sistema.
- Los sistemas no prevén sincronización de labios con el mensaje verbal.

Es por ello que la búsqueda por producir personajes animados creíbles (lo que implica una correcta sincronización de la voz con el movimiento de los labios), ha motivado a diversos grupos a desarrollar soluciones que permitan sincronizar de manera eficiente, en tiempo real y de manera creíble, el audio con los movimientos bucales.

Así, existe en el mercado un número creciente de programas de cómputo y *plug-ins* (módulos externos que extienden las capacidades de un programa) de origen comercial, que ofrecen soluciones parciales al problema de sincronización. Estos sistemas se basan en la correlación entre un archivo de audio y uno de texto (ver Figura 1), en el cual se encuentra la trascripción del discurso del archivo de audio. Esto permite una correcta selección de v*isemas*[3] y fonemas. Sin embargo, ninguno de estos sistemas permite realizar procesos en tiempo real, ya que todos dependen de correlacionar manualmente los visemas/fonemas en función del discurso en el tiempo (Figura 2).

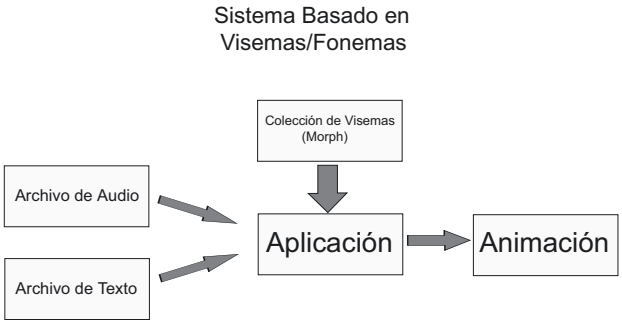
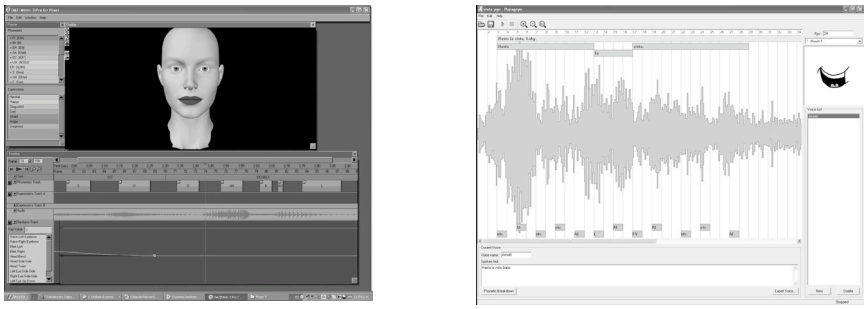


Figura 1



a) *b)*

Figura 2. Sistemas de Lipsync comerciales basados en fonemas y visemas.
a) Mimic 2.0 elaborado para figuras en 3D b) Papagayo 1.0 Elaborado para figuras en 2D.

Así, a pesar de ser una solución en boga, resulta insuficiente cuando de lo que se trata es de producir movimientos faciales de manera automática y en tiempo real. La solución actual, además de requerir grandes cantidades de horas-hombre para conseguir la correspondencia entre los visemas y el archivo de audio, no es capaz de adaptarse fácilmente a nuevos textos.

METAS

Por lo anterior en este trabajo[4], nos proponemos atender esta necesidad y establecemos dos metas para seguir:

- Definir un algoritmo que permita la asociación automática de los movimientos de labios a un archivo de audio.
- Definir un proceso que permita añadir al sistema nuevos archivos audio, desde su registro hasta la integración en el funcionamiento del sistema.

A continuación se describe la manera en la cual hemos abordado las metas presentadas. Primero se presentará el algoritmo para asociar movimientos de labios con el archivo audio (secciones 1-2) y luego el proceso de integración a un sistema (sección 3). En la sección 4 se hace una discusión sobre los resultados, mientras que en la sección 5 se enumeran las líneas de investigación para seguir en el futuro.

EL PROBLEMA

En el marco del sistema Enciclomedia, desarrollado por la Secretaría de Educación Pública de México y el Instituto Latinoamericano de Comunicación Educativa (ILCE),

se diseñaron e implementaron un grupo de compañeros virtuales con movimientos preprogramados, los cuales pueden aparecer directamente o cuando un medio (video, lectura, simulación, etc.) termina, lanzando una pregunta relacionada con el/los objetos de aprendizaje del material que se presentó en una lección de los Libros de Texto Gratuitos. Con el fin de mantener la atención del grupo, dicha pregunta debe ocurrir de manera verbal, asociada a manifestaciones corporales preprogramadas, lo cual implica un sistema de sincronización de audio relacionado con las partes móviles de la cara del Avatar. Las preguntas varían dependiendo de cada sección del software y se requiere la posibilidad de añadir nuevos archivos con preguntas, por lo que una solución basada en visemas resulta inoperante.

En la propuesta que aquí se presenta, el primer paso del trabajo es el análisis de un archivo que contiene el audio de la pregunta a formular. Este análisis permite la identificación automatizada de las partes de habla y las de silencio. Este análisis genera un archivo en el que se sincronizan en una misma línea de tiempo el audio y las diferentes representaciones de la boca. La animación automatizada de la boca en sincronía con el archivo se trata posteriormente en este trabajo.

Aunque simple en su formulación, la puesta en práctica de este paso impone buscar respuestas a varias preguntas relacionadas, por ejemplo:

- ¿Cuál es la influencia del formato del archivo de sonido sobre el proceso de análisis?
- ¿Cómo se puede caracterizar el silencio? Y en consecuencia ¿cómo diferenciar el ruido de fondo del silencio?
- ¿Cómo reducir la granularidad de la información contenida en la muestra?
- Al considerar la parte de animación, hay que definir ¿que resolución de cuadros usar para la animación?
- ¿Cómo elegir un diseño de boca con base a la información de muestreo?

Responder a estas preguntas impone al mismo tiempo determinar el proceso de trabajo, desde la concepción general de como se integra un componente de sincronización a un proyecto más grande, hasta determinar los detalles referentes al archivo de audio por ejemplo; codificación usada, resolución de muestreo, condiciones generales para registrar el archivo, etc.

La hipótesis del trabajo plantea que sí es posible detectar de manera automática zonas de habla y de silencio, y a partir de ello, hacer la animación de la boca en sincronía con el archivo de sonido basándose *solamente* en la información contenida en el muestreo.

Los humanos percibimos los movimientos de la boca y tenemos la expectativa natural que estos correspondan al texto o sonido que se escucha, es decir, la expectativa es de ver visemas asociadas adecuadamente al sonido que se escucha. Sin embargo, la información contenida en un archivo de sonido se refiere a la frecuencia en un momento de muestreo y por lo tanto no retiene ninguna información sobre que se dice en la grabación.

Varias investigaciones confirman la dificultad de la tarea de sincronización sin tener el texto leído disponible. Nosotros, a partir de la hipótesis planteada, mostraremos que es posible hacer una aproximación satisfactoria de los movimientos de la boca, sin la necesidad de llegar al nivel de visemas.

METODOLOGÍA

Análisis del archivo de sonido

Formato del archivo

Los formatos más usuales para archivos de sonido son el WAV y MPEG, de los cuales el más conocido es el MP3. El formato MP3 se volvió muy popular dado que la codificación usada permite reducir drásticamente la dimensión del archivo y mantener al mismo tiempo la calidad del sonido. Por lo tanto, también han proliferado los programas para leer estos archivos. El principio de organización consiste en guardar la información en cuadros (*frames*) y en cada uno de ellos se tienen al principio las especificaciones (*header*). Tal organización permite escuchar pedacitos del archivo de sonido, dado que toda la información necesaria para tocarla está contenida en el mismo cuadro. Dicha organización permite que los cuadros tengan características diferentes, es decir muestreo a frecuencias diferentes de tiempo o codificación del audio en diferente calidad, etc.

El formato WAV está estandarizado, lo que representa una gran ventaja cuando se trata de analizar los datos contenidos. El principio de organización es diferente del MPEG, aquí tenemos una organización global de la información. El archivo de sonido tiene una parte descriptiva (similar a los archivos MP3, pero la descripción es global) que contiene información sobre todo el archivo de sonido. Esta parte del archivo está seguida por los datos correspondientes al muestreo. Por la organización del archivo tenemos datos correspondientes a momentos de tiempo iguales. La gran desventaja de un archivo de sonido en este formato es la limitación del tamaño, dado que en la parte de descripción del formato hay disponibles solamente 4 bytes para guardarlo.

El proceso de análisis debe rastrear el archivo y extraer los datos de muestreo. Hemos efectuado varios experimentos con los dos formatos y hemos llegado a la conclusión que el formato MPEG regresa datos que no corresponden con el sonido (al escuchar). En la Figura 3 se presenta la comparación entre los resultados del análisis del mismo archivo de sonido, pero en los dos formatos, WAV y MP3. Se puede observar que hay diferencias fuertes aunque estamos hablando del mismo archivo. Frente a tal situación hemos decidido usar para el análisis el formato WAV.

Archivo WAV

0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,3100,6968,-1270,-1236,1901,2794,6090,3583,3286,-2791,-1225,8243,8630,-370,-219,-1162,30,0,0,0,0,0,0,0,0,0,-3,4635,-5663,-983,1598,2322,10471,-680,-389,-419,-10675,-14,-13269,6239,5948,99,1407,-34,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0

Archivo MP3

1,-1,0,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1025,-1,-1,0,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1025,-1,-1,0,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1025,-1,-1,0,-1,-5754,-29546,-5066,451,-30288,6151,25799,11056,-19292,22319,3372,32445,1267,17661,26380,-1025,11448,-5790,0,-17758,-27779,-1104,2944,-6384,-2139,22037,- 1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1025,-1,-1,0,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1025,-1,-1,0,-1,-1,-1,-1,-1,-1,-1,-1,-1,0,0,0,0,0,8224

Figura 3. Comparación de lectura.
Ambos análisis corresponden a la misma grabación.

Ruido de fondo

Para hacer un análisis adecuado se necesita saber sobre la presencia del ruido de fondo. Al considerar que el ruido puede ser interpretado como dato real, es necesario saber de su presencia. Una solución consiste en incorporar un segundo de *silencio* al principio de la grabación para registrar el ruido inherente del micrófono y al ambiente y, posteriormente a la grabación quitar el ruido de fondo de los datos de audio. Otra alternativa es realizar las grabaciones en condiciones de estudio, opción que exige una infraestructura adecuada y por lo tanto es poco viable. Además de decidir que es el ruido en un audio, se necesita decidir que se interpreta como silencio. Hemos decidido identificar con silencio los datos que están menores a un umbral que se definió de manera proporcional con los datos presentes en el archivo.

Resolución de los datos

El número de muestras por segundo es un parámetro del formato WAV, por lo tanto hay que leerlo e identificarlo adecuadamente para extraer los datos. A cada segundo vamos a tener un gran número de datos. En caso de tener ruido de fondo, guardado en el primer segundo del archivo, se ajustan los datos leídos para eliminar el ruido de fondo. Una vez obtenidos las muestras nos encontramos con el problema de reducir la granularidad de la información, es decir extraer un número menor de datos que se puedan considerar como representativos para un periodo de tiempo deseado (un segundo o fracciones de segundo).

El problema se torna complejo, porque hay que determinar criterios para comprimir datos. Esto lo convierte en un problema que pertenece al área de compresión y codificación de datos.

Vamos a ilustrar el problema con el siguiente ejemplo: supongamos que tenemos 25 datos correspondientes a un cuarto de segundo y queremos asignar un solo valor que contenga la *esencia* de los 25. Consideramos los siguientes 4 casos (para simplificar vamos a usar solamente 1 y 0's, donde 1 marca la presencia de un dato y 0 representa *silencio*).

a. 0001110001110001110001111
b. 0000000000000111111111111
d. 1111111111100000000000000
d. 1111111111110000000000000

El propósito es asignar solamente un valor a estas cadenas. ¿Qué criterios se podrían usar? En principio se podría pensar en decidir con base en el promedio de 1's, si es mayor a la 0.5 entonces el valor asignado sea 1, en caso contrario 0. Si aplicamos este criterio solamente en el caso *d* tendremos 1. Sin embargo, es evidente que estamos perdiendo una información valiosa, dado que en todos los casos el promedio es muy cercano a 0.5.

Otra idea sería la de decidir dependiendo de la secuencia de 1's y 0's tomando un criterio más complejo con algún promedio ponderado (según un peso asociado de manera subjetiva con la posición de aparición en la cadena u otro criterio). Nuestra conclusión es que no hay garantía de que los criterios determinados funcionen satisfactoriamente en todas la situaciones. Este resultado tiene un gran impacto sobre el proceso de trabajo, porque significa que podemos buscar solamente

soluciones adecuadas a una situación en particular y no globales. Para tal propósito vamos a tener que analizar y aprovechar las particularidades del contexto en el cual nos encontramos y necesitamos evaluar el costo de un error de evaluación en el desempeño de la animación final.

Sincronización bucal

Resolución de cuadros

La respuesta a la pregunta ¿cuántos cuadros habrá que usar para tener una dinámica de la boca percibida como natural? surge de trabajos de animación. En el presente trabajo se optó por 12 cuadros por minuto. Esta resolución asegura una percepción natural de los movimientos de boca y facilita la reducción de información por segundo del archivo audio. Al analizar se observa que una resolución (más baja de cuadros por minuto) hace que la animación se vea demasiado lenta, además que un error (de asignación del diseño de boca con la fracción del audio: es decir, asignar *silencio* cuando hay habla) es molesto.

Elección de la secuencia de animación

Una vez determinada la resolución por minuto de los cuadros y el algoritmo de reducción de granularidad de los datos de audio hay que determinar la manera de asociarlos con secuencias animadas. En primer lugar se diseñaron 17 animaciones cortas para constituir el conjunto de *bocas* posibles. Estos diseños van desde la boca cerrada hasta una boca abierta por completo (ver figura 5). El fragmento del archivo audio se asocia con un archivo de animación (que contiene la animación de la boca) con base en los valores relativos del fragmento con respecto al archivo audio total. Los intervalos de valores designados a cada animación (es decir, la correspondencia entre archivo audio y archivo de animación) son proporcionales con el valor del audio correspondiente a una unidad de tiempo (aquí la 1/12 parte de un segundo). Con la proporcionalidad se evite el uso excesivo de bocas muy abiertas, además de excluir cambios bruscos en la animación.

IMPLEMENTACIÓN

En la figura 4 se presenta un esquema del sistema implementado para leer archivos WAV y cuya salida es una animación bucal en Flash.

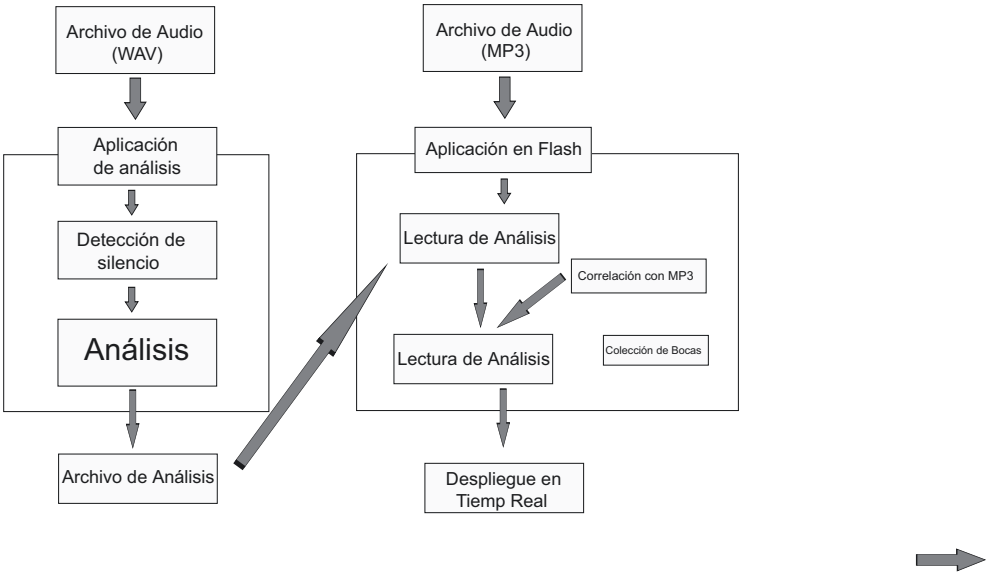


Figura 4. Diagrama del sistema completo.
Panel Izquierdo, Unidad de Análisis. Panel derecho Unidad de Despliegue Gráfico.

En el panel izquierdo se diagrama la secuencia de trabajo de la Unidad de Análisis, la cual abre el archivo WAV, determina el silencio en el primer segundo, determina la velocidad de muestreo del análisis y produce un archivo de texto similar al panel superior de la figura 3, el archivo entonces es guardado en la base de datos de las preguntas junto con la versión en MP3 del audio, de tal suerte que al ser llamado por la Unidad de Despliegue Gráfico correspondan el audio y el análisis.

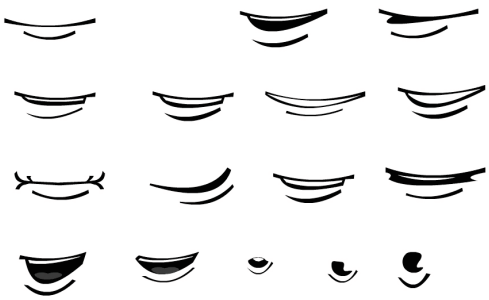


Figura 5. Colección de pseudovisemas.
Primera a la izquierda se muestra "boca cerrada", que se asocia con el silencio.

En el panel derecho se presenta la secuencia de pasos de la Unidad de Despliegue Gráfico, donde es leído el archivo de Análisis y correlacionado con el MP3 del audio, en cada paso de muestreo, se intercambian las bocas asumiendo que 0 es la que corresponde a boca cerrada, éste módulo contiene la colección de las bocas en acción (ver figura 5) y las presenta conforme se leen sincronizadas con el archivo de audio, permitiendo el despliegue gráfico en tiempo real.

DISCUSIÓN

Independientemente de los problemas inherentes a la granuralidad del muestreo, el sistema actualmente presenta una solución satisfactoria a los requerimientos, dado que se presenta la animación de la boca en el contexto del avatar, el cual a su vez se encuentra animado.

La sincronización obtenida recuerda las caricaturas comerciales de los sesenta, dando una credibilidad mínima necesaria para que el usuario no sea distraído por una mala correlación de visemas/fonemas.

Gracias a lo anterior, el sistema permite: a) generar líneas de producción simples que aceleran el proceso de elaboración de las secuencias verbales, y b) que, personal poco entrenado, pueda agregar comunicaciones verbales de manera independiente.

De esta manera el sistema podrá crecer de manera liberado del apoyo de unidades especializadas en la generación de medios, dando a los profesores la posibilidad de traducir las preguntas en español a las lenguas locales donde se utilice la versión 2.0 de Enciclomedia o generar sus propias aproximaciones a los objetos de aprendizaje, dándole libertad de cátedra en este sentido.

TRABAJO A FUTURO

Preguntas pendientes

Aún cuando los resultados del presente trabajo han sido suficientes para la implementación de un sistema de Avatares parlantes en 2D, quedan por resolver tres problemas derivados del sistema de análisis de audio. El primero: cómo decidir cuando se pueden repetir los pseudovisemas, tiene que ver con el tipo de análisis, dado que el análisis actual no permite diferenciar si dos muestras de audio son idénticas o su diferencia es de calidad y no de cantidad de señal. El segundo: cuántos

pseudovisemas necesitamos para tener una variedad que asegure la sensación de naturalidad; y por último como se percibe si hay una falla, en el sentido de tener una boca cerrada en lugar de boca abierta.

Ampliación tecnológica

En futuros trabajos, estaremos obligados a mejorar el análisis del audio con la meta de poder utilizar un micrófono de baja calidad que pueda ser compatible con cualquier máquina.

Sistemas 3D

La meta en el mediano plazo es utilizar una modificación de la unidad de despliegue gráfico en la representación bucal en tiempo real, de avatares que se encuentren en sistemas 3D. De esta manera, se podrán producir comunidades virtuales colaborativas en las que sea posible que los usuarios se comuniquen a través de Avatares *parlantes*, lo cual permitirá relegar las comunicaciones de tipo *sólo texto* a una opción y no la única opción.

NOTAS

1. Avatar es un término de origen sánscrito (avatâra significa *el que desciende*), que en el marco del hinduismo se refiere a la encarnación terrestre de un dios. Por ejemplo, se dice que El Señor Krishna es la octava encarnación (Avatar) de Vishnú.
2. También conocidos como *Agentes sintéticos*, agentes de software que operan en ambientes simulados, como mundos virtuales, o juegos de video. Se enfatizan cualidades como una representación gráfica, credibilidad y personalidad, en lugar de inteligencia o especialización, y pueden jugar papeles en sistemas interactivos para entretenimiento, arte y educación (Isbister y Doyle, 2002).
3. Visema se define a la posición que los labios asumen cuando se produce un determinado fonema.
4. El presente proyecto fue patrocinado parcialmente por el Macroproyecto de Tecnologías para la Universidad de la Información y la Computación Programa Transdisciplinario en Investigación y Desarrollo de la Universidad Nacional Autónoma de México.

REFERENCIAS BIBLIOGRÁFICAS

André, E.; Rist, T.; Muller, J. (1999). Employing AIl methods to control the behavior of animated interface agents. *Applied Artificial Intelligence*, 13, 415-448.

Conati, C.; Zhao, X. (2004). Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game, en: Vanderdonckt, J. y otros *International Conference on Intelligent User Interfaces, Proceedings of the 9th international conference on Intelligent User Interfaces*, *Funchal,* Madeira, Portugal, ACM Press, 6-13.

Hewitt, C. (1997). Viewing control structures as patterns of passing messages. *Journal of Artificial Intelligence*, 8(3), 323-364.

Isbister, I.; Doyle, P. (2002). Design and Evaluation of Embodied Conversational Agents: A Proposed Taxonomy en *The First International Joint Conference on Autonomous Agents & Multi-Agent Systems*, Bologna, Italy.

Lester, J. C.; Converse, S. A.; Stone, B. A.; Kahler, S. E.; Barlow, S. T. (1997a). Animated pedagogical agents and problem solving effectiveness: A large scale empirical evaluation en *Proceedings of the Eighth World Conference on Artificial Intelligence in Education*, IOS Press, 23-30.

Lester, J. C.; Converse, S. A.; Kahler, S. E.; Barlow, S. T.; Stone, B. A.; Bhogal, R. (1997b). The persona effect: Affective impact of animated pedagogical agents, in: *Proceedings of the Conference on Human Factors in Computing Systems*, Atlanta, 359-366.

Lester, J. C.; Voerman, J. R.; Towns, S. G.; Callaway, C. B. (1999). Deictic believability: Coordinating gesture, locomotion and speech in life-like pedagogical agents. *Applied Artificial Intelligence*, 13, 383-414.

Lester, J. C.; Stone, B. A.; Stelling, G. D. (1999). Life-like pedagogical agents for mixed-initiative problem solving in constructivist learning environments. *User Modeling and User-Adapted Interaction*, 9, 1-44.

Shaw, E.; Ganeshan, R.; Johnston, W. L.; Millar, D. (1999). Building a case for agent-assisted learning as a catalyst for curriculum reform in medical education en *Proceedings of the Ninth World Conference on Artificial Intelligence in Education*, Le Mans, IOS Press, 70-79.

PERFIL ACADÉMICO Y PROFESIONAL DE LOS AUTORES

Fernando Gamboa Rodríguez. Líneas de investigación: Estudio de Interacción Humano-Máquina, Desarrollo de aplicaciones educativas interactivas centradas en el usuario.

E-mail: fernando.gamboa@ccadet.unam.mx

M. C. Ildikó Pelczer. Maestría en Ciencias de la Computación, actualmente inscrita en el Programa de Doctorado de las Ciencias de la Computación.

Líneas de investigación: inteligencia artificial, enseñanza de las matemáticas, avatares, modelación del usuario, entornos virtuales.
E-mail: IPelczer@iingen.unam.mx

Francisco Cabiedes**.** Líneas de desarrollo: desarrollo de sistemas Avatar-Agente, entornos virtuales, estereoscopía, visualización educativa.
E-mail: caviedes@aleph.cinstrum.unam.mx

DIRECCIÓN DE LOS AUTORES

Universidad Nacional Autonóma de México
Circuito Interior s/n Coyoacán
04510 México, D.F. México

M. en C. Steve Rodríguez Rodríguez. Líneas de desarrollo: Educación asistida por computadora, inteligencia artificial, ciencia cognitiva, recuperación de información, desarrollo de Enciclomedia.
E-mail: steverdz@hotmail.com

D. G. Javier Bretón Palomo. Líneas de desarrollo: Desarrollo de interfaces gráficas para Enciclomedia.
E-mail: javo1doble0@hotmail.com

DIRECCIÓN DE LOS AUTORES

Dirección de Investigación e Innovación,
Enciclomedia,
Instituto Latinoamericano de la Comunicación Educativa
Periférico Sur 4118, Torre Zafiro 1, 7mo Piso
01900 México, D.F. México

Fecha de recepción del artículo: 23/10/06
Fecha de aceptación del artículo: 07/03/07