

# **Animación Facial y Corporal de Avatares 3D a partir de la Edición e Interpretación de Lenguajes de Marcas.**

M.P Carretero, D. Oyarzun , I. Aizpurua, A. Ortiz

Departamento de Edutainment y Turismo.  
Asociación VICOMTech, Paseo Mikeletegi 57, 20009 San Sebastián, España  
Tlf: 943 30 92 30 Fax: 943 93 93.  
e-mail: mcarretero, doyarzun, iaizpurua, aortiz@vicomtech.es

## **Resumen**

Los personajes virtuales se están utilizando cada vez más en las interfaces de usuario para mejorar la comunicación persona-máquina. Por este motivo, surge la necesidad de definir el comportamiento y la apariencia de estos personajes de una manera intuitiva y visual a través de editores gráficos. Tras realizar un análisis de los principales lenguajes de marcas para definir animación de personajes virtuales y teniendo en cuenta las técnicas de animación facial y corporal utilizadas para la animación de nuestros personajes, nos hemos decidido por desarrollar un editor gráfico para el lenguaje de marcas VHML. En este artículo se presenta el proceso de desarrollo del editor gráfico basado en este lenguaje, así como el de su interprete que consigue la animación en tiempo real.

**Palabras clave:** VHML, Animación en Tiempo Real, Avatares, Personajes Virtuales.

## **1 Introducción**

En los últimos años está creciendo el uso de aparatos electrónicos como el PC, el móvil o el PDA, ya que además de facilitar ciertas tareas de la vida laboral también cumplen funciones de entretenimiento o aprendizaje. Este crecimiento no sólo se debe a que el precio de estos productos está bajando, sino a que están mejorando continuamente y su uso resulta cada vez más fácil y atractivo. Sin embargo, sigue existiendo la necesidad de conseguir una comunicación más personal y natural,

basada en la comunicación real entre los humanos. Aparece un nuevo campo de investigación que centra su trabajo en el área de Interfaces Conversacionales de Usuario, incluyendo los Asistentes Virtuales.

Estos asistentes, llamados avatares, son personajes virtuales cuyo objetivo principal es hacer que la comunicación entre el usuario y la máquina sea más natural e interactiva. Estos cumplen una serie de funciones en el sistema, como la de asistencia o guía a través de la aplicación o simplemente la de presentar la información. La apariencia del personaje puede ser de tipo antropomórfico o de tipo *cartoon*, así como 2D o 3D dependiendo del dispositivo en el que se esté ejecutando (PC, PDA, móvil, televisión digital, etc). La principal ventaja de utilizar avatares en las interfaces, es la ilusión generada al usuario de la existencia de un personaje con el cual puede interactuar como si se tratase de un personaje real, lo que no ayuda sólo en aspectos técnicos de la interacción, sino también en los psicológicos.

Esta ilusión se consigue dando al avatar la posibilidad de expresar emociones, realizar gestos faciales y corporales. Una de las maneras más eficientes de programar en el avatar estas cualidades es a través de texto etiquetado, que contendrá, además del texto que el avatar deberá reproducir, una serie de etiquetas que definirán las emociones, sus intensidades y los gestos faciales o corporales que deberá reproducir el avatar en el momento en que se le indique.

El uso de avatares está extendido a todo tipo de entornos, por este motivo la definición de su comportamiento no puede estar limitada a expertos programadores. Entonces, surge la necesidad de facilitar la tarea de definir las animaciones de personajes virtuales.

Un estudio de la tecnología actual previo al desarrollo de las técnicas descritas en este artículo mostró la total ausencia de editores gráficos que permitieran el etiquetado automático de animaciones para personajes virtuales. Después de realizar un análisis de los principales lenguajes de marcas, expuesto en el capítulo 2 y teniendo en cuenta nuestras técnicas de animación facial y corporal explicadas en el capítulo 3, nos hemos decidido por la implementación de un editor basado en VHML (Virtual Human Markup Language) [1]; ya que, además de permitir una completa definición de la animación de los personajes virtuales, se ha convertido en un estándar que usan proyectos como MetaFace [2] y The Mentor System [3]. El diseño e implementación de este editor se explica en el capítulo 4. Por último, en el capítulo 5, se muestra el motor para interpretar cada etiqueta vhtml y ejecutarlas sobre el avatar utilizando las técnicas de animación comentadas anteriormente.

## 2 Análisis de lenguajes de marcas

La necesidad de etiquetar las animaciones de los avatares nos ha llevado a realizar un análisis de los lenguajes existentes con el objetivo de utilizar un estándar capaz de etiquetar animaciones faciales y corporales que se pudiera adaptar a nuestro motor de animación. La tabla 1 se muestra la comparativa hecha entre los lenguajes de marcado estudiados.

Language	Emociones	Anim. Facial	Anim. Corporal	Diálogo	MPEG4	Proyectos
HumanML [5]	Sí	Sí	Sí	Sí	-	-
CML [6]	Sí	Sí	Sí	-	Sí	SAFIRA
AML [7]	Sí	Sí	Sí	-	Sí	SoNG
APML [8]	Sí	Sí	No	-	-	Greta
DAMSL [9]	-	-	-	Sí	-	-
GESTYLE[10]	Sí	Sí	Sí	-	-	-
MPML [11]	Sí	Sí	No (MPML-VR)	-	-	Agentes Telefonía móvil
MURML [12]	-	Sí	Sí	No	-	Max
VoiceXML [13]	No	No	No	Si	-	-
RRL [14]	Sí	Sí	Sí	Sí	-	NECA
AIML [15]	-	-	-	Sí	-	A.L.I.C.E.
VHML [16]	Sí	Sí	Sí	Sí	Sí	InterFace MetaFace MentorSystem
XSTEP [17]	No	No	Sí	-	-	-

*Tabla 1. Análisis de los principales lenguajes de marcas estudiados.*

Como se puede comprobar en la Tabla 1, AML, CML, RRL y VHML son los únicos que cumplen casi todos los requisitos y permiten la expresión de emociones, animación facial y corporal. Sin embargo, descartamos RRL, pues no cumple con el estándar MPEG-4, requisito que consideramos esencial para integrar los avatares

en dispositivos de telecomunicaciones como la TV Digital, incorporando componentes multimedia con una rápida transmisión. CML también lo descartamos, pues parece que no se va a convertir en un estándar.

VRLab [4] considera a AML y a VHML como estándares para avatares. Estos dos lenguajes cumplen con los objetivos deseados. Permiten la representación de emociones, animación facial y corporal, y los dos cumplen con el estándar MPEG-4. Para tomar la decisión final, como se puede ver en la Tabla 2, se han analizado los dos lenguajes por separado, comparándolos entre sí y viendo qué etiquetas están definidas para cada elemento a animar. Los dos tienen sub-lenguajes para la representación de las animaciones faciales y corporales, y etiquetas para configurar a los avatares que van a ser utilizados.

	<b>VHML</b>	<b>AML</b>
<b>Emociones</b>	EML	-
<b>Gestos</b>	GML	-
<b>Animación facial</b>	FAML	AFML
<b>Animación Corporal</b>	BAML	ABML
<b>Personajes</b>	<person>	<AML face ide....>
<b>Voz</b>	<voice...>	<TTS>

*Tabla 2. Comparativa entre VHML y AML*

Finalmente se ha decidido utilizar como lenguaje de etiquetado el lenguaje de marcas VHML [1]. El primer motivo es que es capaz de englobar todos los aspectos de la interacción persona máquina a través de personajes virtuales. VHML cumple todos los requisitos definidos para ser integrado en nuestro motor de animación ya que define etiquetas para: animación facial, animación corporal, producción texto-habla y representación emocional. Así mismo, permite también definir un gestor del diálogo de interacción y controlar componentes hypermedia y multimedia. Otro de los motivos para su utilización ha sido que se trata de un estándar y que su especificación es pública.

### **3 Motor de Animación**

Para conseguir una animación natural del avatar es imprescindible producir una animación tanto facial como corporal realista. En nuestro caso hemos implementado un motor que permite la animación de ambas partes. Por un lado, consideramos que la expresividad facial está producida tanto por el grado

emocional que producen las facciones del avatar como por el movimiento de los labios al vocalizar sincronizadamente las frases que debe reproducir. Las partes faciales que son habitualmente animadas en el avatar son: la cabeza, los labios, los ojos, las pupilas, las cejas y las pestañas.

Algunas de las animaciones, como enfatizar, asentir o negar, simplemente se realizan generando una deformación individual más una traslación o desplazamiento en una determinada trayectoria, como es el caso de las pupilas o el movimiento de la cabeza. Otras, como las de los labios, sonreír o guiñar un ojo se realizan utilizando técnicas de *Morphing* [18]. En [19] se puede encontrar una explicación más detallada de nuestro motor de animación.

Por otro lado, para la animación realista del cuerpo de un personaje virtual se ha considerado necesaria la integración de dos técnicas:

- Animaciones predefinidas: Para el tipo de animaciones como andar o correr se han utilizado los *routes* de VRML ya que son animaciones recursivas que pueden estar predefinidas sin influir en el grado de interactividad de la aplicación.
- Animaciones basadas en cinemática inversa: Los movimientos de los brazos, como señalar hacia un punto concreto o realizar algún movimiento con las piernas se han implementado siguiendo técnicas de cinemática inversa, ya que en este tipo de movimientos las animaciones pregrabadas no son válidas si se desea un alto grado de interactividad. Un ejemplo del movimiento del avatar reproduciendo este tipo de animación se muestra en la Figura 1.

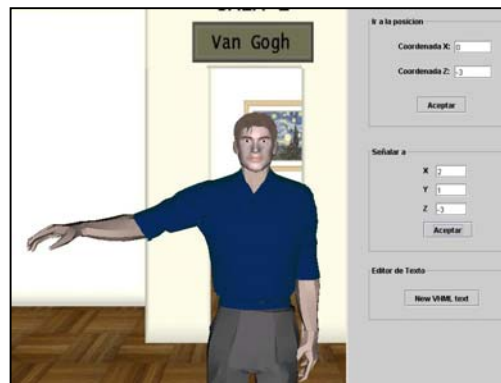


Figura 1. Ejemplo del avatar señalando hacia el punto que se le ha indicado en la ventana de usuario.

El lector puede encontrar una explicación detallada de las bases matemáticas del módulo de cinemática inversa en [19].

## 4 Editor VHML

El hecho de que los avatares aparezcan cada vez más en aplicaciones cotidianas y que los usuarios no especializados los utilicen cada vez más, crea la necesidad de proporcionar medios para diseñar movimientos y características de avatares de forma sencilla. Es decir, que cualquier usuario pueda indicar qué emoción tiene un avatar en un momento determinado, los gestos y movimientos que debe hacer, sin necesidad de conocer el lenguaje de etiquetado. Así, surge como solución el crear un editor gráfico para el etiquetado del texto del avatar.

A través de este editor el usuario podrá definir la animación facial y corporal del avatar sin necesidad de ningún conocimiento de lenguajes de animación 3D. Una de las principales motivaciones de su desarrollo es que no se ha encontrado en el mercado ningún editor que permita realizar esta tarea. Otro aspecto que hace práctica la implementación del editor es la transmisión de las animaciones faciales y corporales a través de la web en un formato de texto y mediante un estándar que haga posible su lectura a un amplio número de usuarios. Con el editor el usuario podrá definir fácilmente la animación del avatar en un lenguaje basado en XML guardándolo en ficheros de bajo peso para su rápida transmisión a través de la red. El editor ofrece la posibilidad de elegir el avatar deseado y configurarlo a su gusto a través de un editor de avatares integrado, de insertar / eliminar emociones permitidas por VHML, gestos definidos por VHML, animaciones faciales definidas por VHML, animaciones corporales definidas por VHML, de generar un fichero (\*.vhml) que contenga el texto etiquetado correctamente y de abrir ficheros generados previamente.

Como se ha mencionado anteriormente, se quiere conseguir una comunicación persona-máquina lo más realista posible. Las animaciones grabadas previamente hacen que la interacción muchas veces no sea natural. Más adelante se verá que el editor de textos etiquetados permite generar animaciones en tiempo real de manera sencilla. Esta es otra de las razones por las que es necesaria la implementación del editor de etiquetado.

### 4.1 Diseño e implementación.

La Figura 2 muestra la interfaz que se ha implementado para el editor. Ha sido diseñada de manera que la definición de la animación pueda resultar intuitiva y natural. Como se puede ver, hay cinco partes diferenciadas en la interfaz. En el área de texto se van señalando con iconos el personaje o personajes que se hayan elegido con los botones definidos para seleccionarlos, también se indicarán con

iconos los gestos y animaciones faciales o corporales que se deseen. Las emociones se diferenciarán mediante diversos colores en el texto. Esta implementación permite al usuario conocer en todo momento cuáles son las cualidades que se le ha dado al avatar y poder modificarlas o eliminarlas si lo desea. Situando el cursor delante de los iconos y pulsando el botón derecho del ratón aparecerá un menú con las posibles acciones que tenemos ante ese evento, como puede ser eliminar, modificar los atributos o ver el texto asociado a ese evento. Además, desde esta aplicación, se puede acceder a un editor de avatares donde el usuario puede definir la apariencia que quiere que tenga el personaje del que va a definir su animación.

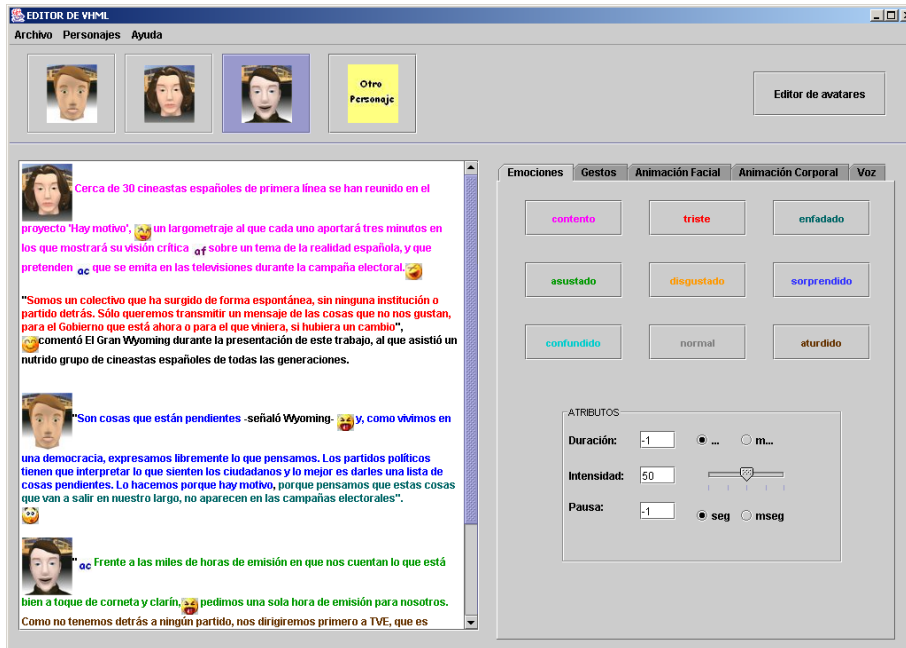


Figura 2. Interfaz del editor VHML

Se ha diseñado la interfaz de esta manera por que resulta intuitiva al usuario para indicar la animación del avatar. Todas las instrucciones que pueden darse al personaje están accesibles visualmente y organizadas de una forma coherente, por emociones, gestos, animaciones faciales y animaciones corporales. Dentro de las animaciones faciales y corporales, los gestos que pueden hacerse se dividen por el órgano que les corresponde. De esta manera se consigue que esté organizado de una forma que resulta sencilla la definición de la animación.

Además, se puede acceder a un manual de usuario desde la barra de herramientas en el que se pueden consultar todos los aspectos necesarios para definir las animaciones utilizando el editor de VHML.

VHML consta de otros sub-lenguajes para el etiquetado del texto:

- EML para emociones
- GML para los gestos
- FAML para animación facial
- BAML para animación corporal

Para la implementación del editor de VHML, se va a tener por cada uno de estos sub-lenguajes, una estructura que permita la representación y almacenamiento de cada uno de los eventos introducidos de cada uno. Estas estructuras almacenarán el *tag* del evento seleccionado y los atributos asociados a él. En el momento de obtener el archivo de salida, la aplicación transforma toda esta información a la semántica de etiquetado definida por VHML. Este modo de actuación permite incorporar nuevos módulos que etiqueten la información almacenada en las estructuras de datos en otros lenguajes de marcas, sin necesidad de modificar los módulos de almacenamiento y definición de la animación.

## 5 Interprete VHML

Para interpretar las etiquetas VHML se ha desarrollado un motor capaz de transformar el fichero definido en una animación automática utilizando las técnicas de animación explicadas en el capítulo 3. Para la primera versión del interprete se han implementado las etiquetas relacionadas con la cabeza del avatar, es decir, las relativas a emociones, gestos y animación facial. Los resultados obtenidos han sido satisfactorios por lo que el intérprete se está extendiendo a etiquetas relacionadas con las animaciones corporales. El módulo intérprete y su relación con el editor VHML y el motor de animación se puede observar en la Figura 3.

Por un lado, a través del editor de avatares integrado en el editor VHML, el usuario define la apariencia que desea que tenga el avatar. Puede definir tantos avatares como desee y la configuración de estos será almacenada en una base de datos para que, cuando se vaya a etiquetar un texto, se pueda elegir el avatar que lo reproducirá. Por otro lado, el editor tiene como salida un fichero \*.vhml que contiene el texto etiquetado con las emociones, gestos y animación facial que el usuario desea que realice el avatar a lo largo de la secuencia de animación. La salida de este editor define la animación del avatar a nivel de texto.

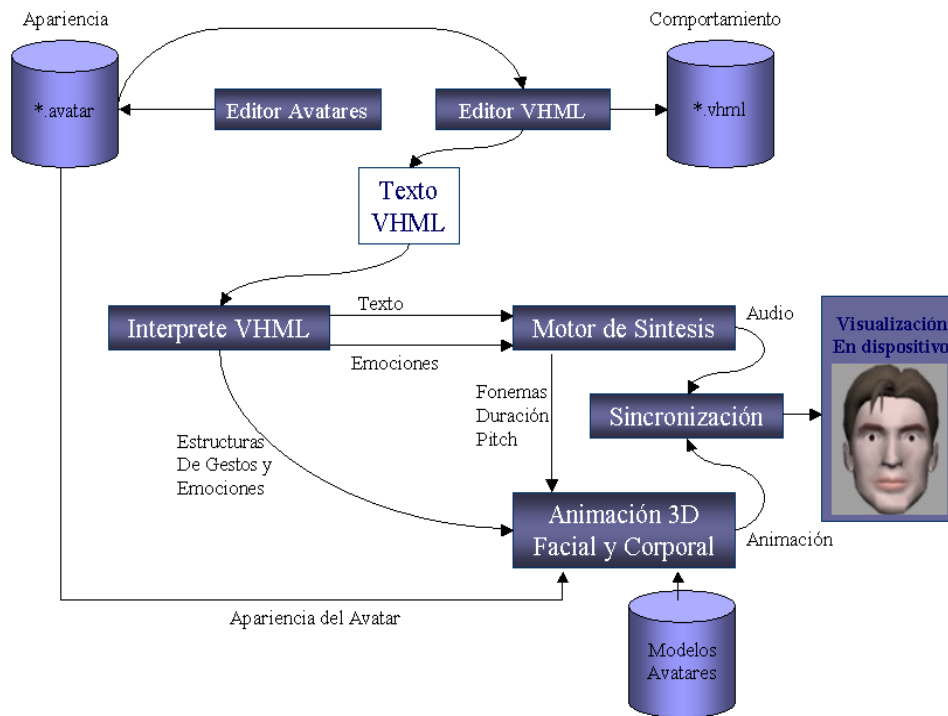


Figura 3 Editor e Interprete

Este fichero es enviado al módulo Interprete de VHML que interpreta el texto etiquetado y extrae la información de cada gesto, emoción y animación facial que se pasan al módulo de animación para controlar la expresión de la cara. El texto a pronunciar, así como las emociones y los eventos relacionados con ellas se pasan también al sintetizador de voz para producir el archivo de audio.

Por un lado, para reproducir la salida gráfica de la animación de los labios se necesitan algunos de los parámetros del sintetizador de voz con el fin de sincronizar los gestos faciales de habla. Principalmente van a ser la frecuencia fundamental o *pitch* y la duración de cada fonema. Con estos datos, se asociará cada fonema a su visema correspondiente -equivalente visual del fonema- mediante técnicas avanzadas de *morphing* [19]. Por otro lado, con la información recibida del interprete VHML el módulo de animación utiliza las técnicas explicadas en el capítulo 3 para convertir cada etiqueta en su expresión facial correspondiente.

## 6 Conclusiones y perspectivas futuras

Se ha desarrollado una herramienta que permite al usuario definir el comportamiento de avatares de una manera sencilla y visual. El resultado será el etiquetado del texto que debe reproducir el personaje, con un lenguaje de marcado basado en XML y que está orientado a la personificación de avatares. El comportamiento será definido mediante las emociones, gestos y movimientos faciales y corporales permitidos por VHML, lenguaje escogido después de realizar un análisis de los principales lenguajes de marcas existentes.

En la revisión de la tecnología actual no se han encontrado herramientas similares por lo que consideramos que este trabajo muestra una aplicación innovadora. Además, su diseño ha permitido obtener una herramienta con una interfaz de usuario intuitiva y fácil de utilizar por cualquier usuario.

Como trabajo futuro hay que destacar que, aunque en el área de texto puedan aparecer varios personajes seleccionados, no indica que interactúen entre ellos, es decir que mantengan un diálogo y la animación esté definida a la vez. Esta opción podría llegar a darse, ya que VHML tiene un sub-lenguaje que permite definir diálogos entre avatares, pero las especificaciones todavía no están publicadas. Por el momento, con el editor de VHML se pueden definir en el mismo fichero la animación de varios personajes, pero sin que lleguen a interactuar entre sí al mismo tiempo.

Además, existe una pestaña en el menú donde pueden elegirse las emociones, animaciones, etc, en la que en un futuro se incluirán las acciones que pueden definirse para la voz del avatar. Esto quiere decir que podrá definirse también el comportamiento de la voz del avatar, con lo que se podrá imitar voces o pronunciar palabras de una determinada manera. VHML tiene etiquetas definidas para poder realizar todas estas acciones.

## Referencias

- [1] VHML Standard. Revisado en Febrero 2003; de: [www.vhml.org](http://www.vhml.org)
- [2] S. Beard, D. Reid, "MetaFace and VHML: A First Implementation of the Virtual Human Markup Language", AAMAS workshop: Embodied conversational agents – let's specify and evaluate the!, Bologna Italy. July 2002.
- [3] A. Marriott, B. Shortland-Jones, "The mentor System", AIED03: Artificial Intelligence in Education, Sydney, Australia. July 2003.

[4] VRLab. Revisado en Febrero 2003; de : [http://vrlab.epfl.ch/research/S\\_standards.html](http://vrlab.epfl.ch/research/S_standards.html)

[5] R. Brooks, K. Cagle, "The Web Services Component Model and HumanML ", OASIS/HumanML Technical Comitee, 2002.

[6] Y. Arafa, A. Mamdani, "Scripting Embodied Agents Behaviour with CML: Character Markup Language", Intelligent User Interfaces, Miami, Florida 2003

[7] S. Kshirsagar, N. Magnenat-Thalmann, A. Guye-Vuillème, D. Thalmann, K. Kamyab, E. Mamdani, "Avatar Markup Language", ACM International Conference Proceeding of the workshop of Virtual environments, Barcelona, Spain, 2002

[8] B. De Carolis, F. de Rosis, V. Carofiglio, C. Pelachaud, I. Poggi, "Interactive Information Presentation by an Embodied Animated Agent", International Workshop on Information Presentatation and Natural Multimodal Dialogue, Verona, Italy 2001

[9] P. Paggio, B. Jongejan, "Multimodal Communication in the Virtual Farm of the Staging Project", International Workshop on Information Presentatation and Natural Multimodal Dialogue, Verona, Italy 2001

[10] H. Noot, Z. Ruttkay, "The GESTYLE Language", International Workshop on Gesture and Sign Language based Human-Computer Interaction, Genova, Italy 2003

[11] M. Ishizuka, T. Tsutsui, S. Saeyor, H.Dohi, Y.Zong, H.Prendinger, "MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions", WebNet, San Antonio, Texas, USA 2000

[12] A. Kransted, S. Kopp, I. Wachsmuth, "MURML: A Multimodal Utterance Representation Markup Language for Conversational Agents", Proceedings of the Workshop Embodied Conversational Agents - let's specify and evaluate them, Autonomous Agents & Multi-Agent Systems, Bologna, Italy 2002

[13] VoiceXML Forum: "Voice eXtensible Markup Language, VoiceXML", Revisado en Febrero 2004; en: [www.voicexml.org](http://www.voicexml.org)

[14] The Neca RRL. Revisado en Febrero 2003; en: <http://www.ai.univie.ac.at/NECA/RRL/>

[15] A.L.I.C.E. AI Foundation. Revisado en Febrero 2003, en: <http://alice.sunlitsurf.com/alice/aiml.html> , <http://www.alicebot.org/>

[16] A. Marriot, "VHML – Virtual Human Markup Language", OZCHI Workshop Talking Head Technology, Fremantle, Western Australia, November 2001

[17] Z. Huang, A. Eliens, C. Visser, "XSTEP: A Markup Language for Embodied Agents". 16th International Conference on Computer Animation and Social Agents (CASA 2003), New Brunswick, New Jersey, 2003

[18] M. Alexa, J. Behr, W. Müller, "The Morph Node". Proc. Web3d/VRML 2000, Monterey, CA., pp. 29-34

[19] D.Oyarzun, A.Ortiz, I.Aizpurura, J.Posada, "Asistentes tridimensionales multilingües para entornos educacionales y de turismo virtual", V Jornadas de Informática y Sociedad, JIS 2004