Primeros Pasos Hacia la Especificación Formal de Interacción Multimodal en Escenarios 3D[1]

Héctor Olmedo Rodríguez , David Escudero Mancebo, Arturo González Escribano, César González-Ferreras, Valentín Cardeñoso Payo

ECA-SIMM, Dpto. Informática. Escuela Técnica Superior de Ingeniería Informática
Campus Miguel Delibes s/n. 47011 - VALLADOLID
hector.olmedo@alumnos.uva.es

Resumen. Basándonos en una filosofía de integración de componentes de aplicaciones de interacción multimodal en entornos gráficos 3D, reutilizando lenguajes de marcas ya definidos para describir gráficos, interacción gráfica y vocal en base a la metáfora de película cinematográfica interactiva, se busca definir un lenguaje de marcas para modelar escenas, comportamiento e interacción que sirva de marco común para desarrollar este tipo de aplicaciones.

Palabras clave: Realidad virtual, comportamiento, interacción vocal, interacción gráfica, interacción persona-ordenador, multimodalidad, sistemas de diálogo, avatar

1. Introducción

Los sistemas de Realidad Virtual (RV) incrementan significativamente el potencial de Interacción Hombre Máquina [1]. Sucede lo mismo con los Sistemas de Diálogo (SD), que aportan un canal complementario al canal gráfico como es el canal vocal o sonoro [2]. La integración de ambos campos de investigación, aunque se ve como una evolución natural de ambas tecnologías, no ha sido apenas explotada en sistemas comerciales, y aunque existen prototipos [3], se trata de un ámbito de trabajo por descubrir. La principal razón por la cual no existen apenas soluciones integradoras de RV y SD, es la juventud de estas áreas de trabajo, donde la mayoría de los esfuerzos se han centrado en mejorar de forma separada ambos campos, y no en estudiar las necesidades de interdependencia que se derivan de una propuesta integradora. Aquí se presenta una propuesta que combina RV-SD planteando una plataforma multimodal de desarrollo de mundos virtuales 3D basada en diálogos.

Los ámbitos de la RV y de los SD se caracterizan por una relativa disponibilidad de prototipos realizados en laboratorios de investigación y de algún sistema comercial, que, por lo general, han descuidado la necesidad de ajustarse a algún

[1] Este trabajo ha sido financiado parcialmente por la Consejería de Educación de la Junta de Castilla y León, en el marco del proyecto VA053A05

estándar de desarrollo o de especificación. El estándar en SD es VoiceXML [4] mientras que en RV hay un estándar de definición de escenas X3D [5] evolucionado de VRML [6]. La disponibilidad de estos estándares ha supuesto un marco de referencia para que los desarrolladores adapten sus sistemas, con las consiguientes aportaciones en cuanto a facilidad de uso en lo que se refiere a la definición de escenarios 3D y diálogos y portabilidad de módulos reutilizables. Aquí se presenta un marco de referencia que pretende ser un lenguaje de especificación de mundos 3D con integración de diálogos. La solución que aportamos respeta los estándares disponibles para RV y SD y sirve de vínculo entre ambos mundos, dando una coherencia argumental a la definición de escenas 3D con interacción hablada.

2. Interacción multimodal 3D y lenguajes de marcado

Añadir interacción vocal a los entornos virtuales con interacción gráfica aporta beneficios claros. Gracias a ello podemos emitir comandos manteniendo la libertad de manos y ojos. Además, los usuarios pueden referirse a objetos que no están presentes en la vista actual del mundo virtual, lo que hace que las acciones sean rápidas y su efecto inmediato. Pero existe una dificultad para la aproximación general a fusión multimodal que hace necesaria la definición de una arquitectura reutilizable para construir nuevos sistemas multimodales.

Las tres componentes de la interacción multimodal para entornos 3D son: la **especificación 3D** que básicamente consiste en modelar objetos del entorno virtual que pueden ser estáticos y/o dinámicos; la **interacción gráfica (GUI)** basada en teclado y ratón como la conocemos hasta ahora y que siempre gira en torno al modelo de eventos y en base a espacios de acción o ***action spaces*** [7] que son aproximaciones metafóricas para estructurar los interfaces de usuario tridimensionales; y por último, la **interacción vocal (VUI)** en la que son posibles cuatro metáforas de interacción: proxy o delegado, divinidad, telekinesis o agente interfaz [8]. Elegir la metáfora de interacción vocal adecuada a nuestro mundo es tarea difícil y más especificar un lenguaje que englobe ésta dentro del marco definido.

Existen numerosos lenguajes de marcas para especificar interacción vocal, escenas y comportamiento por separado pero también hay otros lenguajes de marcado que denominaremos híbridos. Ejemplos de lenguajes de marcado para especificar interacción vocal son **VoiceXML** [4], **SALT** [9] y **X+V** [10]. Entre los lenguajes de marcado para especificar escenas tenemos **VRML** [6] y **X3D** [5]. Las limitaciones de estos dos lenguajes para especificar el comportamiento de los elementos integrantes de la escena ha llevado a definir lenguajes de marcado para especificar comportamiento, como por ejemplo **Behavior3D** [11] o **VHML** [12]. Entre los lenguajes de marcas híbridos podemos citar **MIML** [13] que permite integrar información de discurso, gesto y el contexto de una aplicación dada usando una aproximación de parseo o procesamiento sintáctico/semántico y **MPML-VR** que es una extensión de MPML (Multimodal Presentation Markup Language) un lenguaje de marcas diseñado para presentaciones multimodales que usa VRML 2.0 para poder presentar espacios tridimensionales a través de un agente antropomórfico o avatar de aspecto humano [14]. Además de los lenguajes de marcado vistos existen otros que

también buscan definir especificaciones para aplicaciones concretas [15]. Siguiendo esta corriente y con el objetivo de definir un lenguaje de marcas para la especificación de mundos virtuales con interacción multimodal, llegamos a proponer el lenguaje XMMVR que describiremos en el siguiente apartado.

3. El lenguaje XMMVR

El *eXtensible markup language for MultiModal interaction with Virtual Reality worlds* o **XMMVR** es una propuesta de definición de un lenguaje de marcas para definir escena, comportamiento e interacción en el que consideraremos cada mundo o película interactiva como un elemento "xmmvr" basándonos en la metáfora de película cinematográfica. Podríamos decir que es un lenguaje de marcas híbrido porque la idea es utilizar otros lenguajes tales como VXML o X+V para interacción vocal y X3D o VRML para descripción de escena que quedarían embebidos en éste. De esta manera, el procesamiento de los ficheros xml válidos para el DTD de xmmvr permitirá enlazar con los programas y ficheros necesarios para hacer funcionar el mundo especificado. Nuestro sistema xmmvr va a ser dirigido por eventos, por ello habrá que definir una mínima lista de eventos y no va a haber línea de tiempos. Un elemento xmmvr está formado principalmente por el reparto de actores "cast" y la secuencia de escenas "sequence" que marcan el transcurrir del mundo.

El elemento **"cast"** o reparto será el conjunto de actores que intervendrán en el mundo o película "xmmvr" es decir, cada uno de los elementos que tienen una apariencia gráfica especificada por un fichero vrml y un comportamiento que permite una interacción con el usuario. El usuario lo consideraremos como un espectador sin presencia en el mundo pero que interactúa con los actores de éste, por tanto estamos utilizando la metáfora del proxy o delegado de la que hablamos anteriormente. Diremos que un "**actor**" es todo elemento que puede formar parte del mundo definido y que tendrá comportamientos propios que especificaremos con la etiqueta "**behavior**" y un fichero vrml "**vrmlfile**" que lo especifica. Cada comportamiento "**behavior**" se definirá como una pareja de evento y lista de acciones que puede tener cada actor ante una determinada condición "**condition**". Un evento puede ser provocado por el usuario debido a una interacción gráfica "**eventGUI**" o a una interacción vocal "**eventVUI**". Asimismo existen eventos del sistema que sirven para definir el "comportamiento del mundo". La lista de acciones serán una o varias acciones que se generan ante un evento y pueden ser también de carácter gráfico "**actionGUI**", vocal "**actionVUI**" o de interacción con el sistema "**actionSystem**".

Además de lo anterior, tenemos que especificar la secuencia de escenas "**sequence**" en la que tendremos una o más escenas que se presentarán por defecto en el orden en el que se escribieron. En todo mundo "xmmvr" consideraremos que ocurre al menos una escena "**scene**" y que cada escena tendrá a su vez uno o varios fotogramas o "**frame**". Todo fotograma tendrá un escenario asociado descrito por un fichero vrml "**vrmlfile**" y estará formado por ninguna, una o varias instancias de actores que denominaremos "**sprite**" y ninguno, uno o varios comportamientos "**behavior**" asociados.

Con todas estas premisas hemos definido un DTD [16] y estamos desarrollando una aplicación que se basa en un archivo XML válido para dicho DTD. Es un programa interactivo que denominamos "Centro de reciclaje virtual" en el que en la escena principal tenemos sprites de los actores "robot", "truck", "container" previamente definidos en el apartado "cast" con características añadidas y con los que simulamos una selección de residuos en los diferentes contenedores.

4. Conclusiones y trabajo futuro

Con la propuesta de lenguaje XMMVR definida, queremos demostrar que es necesario definir un meta-guión para especificar cualquier mundo virtual que permita interacción multimodal y que éste aporta modularidad obteniendo con ello claridad y aumentando las posibilidades de reutilización y estandarización. Con el fin de comprobar la efectividad del lenguaje propuesto, nos planteamos como trabajo futuro la implementación completa del ejemplo descrito. Por otro lado, esta propuesta sólo considera la metáfora del proxy o delegado por lo que queda pendiente su extensión para poder dar solución a cada una de las metáforas presentadas o a todas globalmente.

Bibliografía y referencias

[1] W. R. Sherman, A. Craig. ***Understanding Virtual Reality: Interface, Application, and Design***, The Morgan Kaufmann Series in Computer Graphics, 2002

[2] D.Dahl. ***Practical Spoken Dialog Systems (Text, Speech and Language Technology)***, Springer, 2004

[3] C. González-Ferreras, A. González Escribano, D. Escudero Mancebo y V. Cardeñoso Payo. ***Incorporación de interacción vocal en mundos virtuales usando VoiceXML***, CEIG, 2004

[4] VoiceXML Forum: "Voice eXtensible Markup Language": http://www.voicexml.org

[5] Extensible 3D (X3D): http//www.web3d.org/x3d.html

[6] Jed Hartman, Josie Wernecke. ***The VRML 2.0 Handbook***, Silicon Graphics, 1994

[7] R. Dachselt. ***Action Spaces - A metaphorical concept to support navigation and interaction in 3D interfaces***; User Guidance in Virtual Environments, Workshop "Usability Centred Design and Evaluation of Virtual 3D Environments", 2000

[8] S. McGlashan, T. Axling. ***Talking to Agents in Virtual Worlds***, UK VR-SIG Conf., 1996

[9] SALT Technical White Paper: http://www.saltforum.org/whitepapers/whitepapers.asp

[10] XHTML+Voice Profile 1.2: http://www.voicexml.org/specs/multimodal/x+v/12/spec.html

[11] R. Dachselt. ***BEHAVIOR3D: An XML-Based Framework for 3D Graphics Behavior***; ACM Web3D, 2003

[12] VHML Standard: http://www.vhml.org

[13] Latoschik, M.E. ***Designing transition networks for multimodal VR-interactions using a markup language***, ICMI, 2002

[14] Naoaki Okazaki y otros. ***An Extension of the Multimodal Presentation Markup Language (MPML) to a Three-Dimensional VRML Space,*** Wiley-Interscience 2005

[15] M.P Carretero y otros. ***Animación Facial y Corporal de Avatares 3D a partir de la edición e interpretación de lenguajes de marcas***, CEIG, 2004

[16] DTD de XMMVR: http://verbo.dcs.fi.uva.es/~holmedo/xmmvr/xmmvr.dtd