

Arquitectura para Interacción Multimodal en los Juegos por Computadora

Pedro C. Santana
Facultad de Telemática
Universidad de Colima
{psantana}@ucol.mx

Resumen

En la actualidad los juegos por computadora requieren de canales de entrada y salida (E/S) sofisticados para facilitar la comunicación con los jugadores. Utilizando diferentes tipos de E/S como el tacto, gestos y movimiento de cabeza se puede lograr una recopilación más eficiente de los datos. En este trabajo se presenta una arquitectura multimodal para juegos por computadora. Esta arquitectura fue probada con un par de prototipos lo cual permitió a los jugadores realizar las tareas del juego de una forma más comfortable.

Abstract

Nowadays computer games require sophisticated input/output (I/O) channels to facilitate communication with players. Using different types of I/O such haptic, gestures and head tracking can achieve a more efficient data collection. This paper presents a multimodal architecture for computer games. This architecture is demonstrated by two prototypes that allowed players to perform the game tasks in a comfortable manner.

1. Introducción

El juego ha sido una actividad de entretenimiento común en muchas culturas a través del tiempo; juegos como el ajedrez tienen cientos de años de historia exitosa. En las décadas recientes, con la llegada de las computadoras personales, surgieron nuevas posibilidades para crear juegos para computadoras fascinantes y emocionantes. Los juegos por computadora pueden aplicar simulaciones complejas para crear un mundo de juego creíble e inmersivo [1].

Aunque la computación abre formas fascinantes de entretenimiento, jugar con los juegos por computadora sigue siendo percibido como una actividad compleja. Por esa razón la interacción natural con los juegos es un problema común al que se enfrentan sus

diseñadores. Una línea de investigación interesante involucra la creación de más formas inmersivas de proveer los datos de entrada a los juegos, donde los movimientos naturales de los jugadores son traducidos en acciones dentro del juego.

Debido a esto, es necesario explorar nuevos paradigmas de interacción como lo son el reconocimiento de gestos, las interfaces háptica, así como las tecnologías de seguimiento en general para proveer técnicas novedosas e interactuar con los juegos de una forma más intuitiva, natural y centrada en el usuario.

Esto puede llevarse a cabo proporcionando al usuario interfaces que le permitan completar las tareas del juego de una forma específica en el contexto actual del uso y permitir al usuario seleccionar cual es la forma más apropiada para interactuar con el juego de forma multimodal.

Los sistemas multimodales procesan entradas de datos combinadas – tales como reconocimiento de voz, reconocimiento de escritura, gestos de las manos y el seguimiento de los movimientos (del cuerpo, la cabeza y la vista) de forma coordinada con el sistema multimedia de salida. Estos sistemas representan una nueva directriz para la interacción humano-computadora que surge a partir de nuevas tecnologías de entrada y salida que están apareciendo actualmente [2].

De hecho, los sistemas multimodales permiten a los jugadores emplear objetos físicos junto con dibujos, gestos y otras modalidades de entrada para interactuar con la información y con otros jugadores.

Este trabajo de investigación presenta el diseño de una arquitectura de interacción multimodal para el diseño de juegos para computadora, los cuales permitan a los jugadores seleccionar el modo de control de acuerdo a sus preferencias y exigencias del juego.

Este trabajo está organizado de la siguiente manera: la sección 2 contiene la descripción de la arquitectura multimodal para juegos de computadora. La implementación de los prototipos se detalla en la

sección 3, y finalmente, las conclusiones se encuentran en la sección 4.

2. Arquitectura multimodal

Con el fin de lograr la funcionalidad del sistema, se está proponiendo una arquitectura multi-capas. A continuación, se describe dicha arquitectura (ver figura 1).

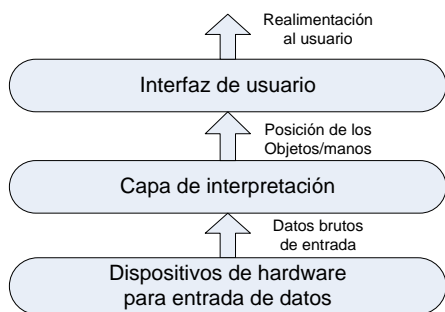


Figura 1 – Arquitectura multimodal

La capa inferior es la capa de hardware, la cual genera los datos brutos de entrada; esto por medio de los diferentes dispositivos de entrada disponibles.

Después de eso, la información es interpretada por la capa de interpretación, la cual traduce los diferentes tipos de datos.

Al final se encuentra la capa de la interfaz de usuario, esta capa genera la parte visible del juego para el jugador. Y recibe los eventos de la capa de interpretación.

3. Implementación

El objetivo principal de este trabajo es lograr la interacción transparente entre los jugadores y el juego por lo que en esta sección se describe el diseño del sistema, la implementación del prototipo y la evaluación de dicho prototipo.

3.1. Diseño del sistema

Para lograr el objetivo, como se ilustra en la figura 2, se han integrado 4 tipos diferentes de interacciones multimodales: Reconocimiento de pluma (*stylus*), reconocimiento por visión, reconocimiento de gestos y seguimiento de cabeza. Los cuales se describen a continuación

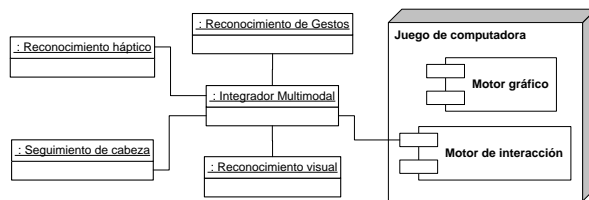


Figura 2. Arquitectura del software.

3.1.1. Motor gráfico. El *framework* XNA de Microsoft [3] fue el seleccionado para hacer más fácil el trabajo con el renderizado de gráficos, implementación de la física, detección de colisiones y todas las funcionalidades base que son utilizadas típicamente en el desarrollo de juegos por computadora.

3.1.2. Motor de interacción. El motor de interacción es el responsable de inicializar los motores de reconocimiento y seguimiento, así como de reconstruir el integrador multimodal (una colección de máquinas de estado finitas).

3.1.3. Integrador multimodal. El integrador multimodal recibe las señales de entrada de los dispositivos externos y construye un significado para el juego. También, mantiene la referencia al estado actual de cada máquina de estado finita.

3.1.4. Reconocimiento de pluma. Para el desarrollo de esta arquitectura de software se utilizó el sistema *DUO digital pen* [4] para reconocer la entrada dada por una pluma. Dicho sistema ofrece la posibilidad de combinar el uso de una pluma ordinaria dentro de una superficie como si se tratara de un ratón para computadora. De esta forma, es posible contar con las funcionalidades de una tableta.

Dadas estas características, el juego toma sus ventajas para controlar al personaje con una pluma y a la vez se reconoce el evento *tap* para disparar acciones en el juego.

3.1.5. Reconocimiento de gestos. El reconocimiento de gestos se desarrolló por medio de un sistema de código abierto con tecnología para detección de sombras llamado *Community core vision* [5]. Dicho sistema trabaja con una cámara web bajo una superficie opaca que permite que la luz se filtre para reconocer los gestos.

El usuario interactúa con el juego moviendo su mano sobre la superficie opaca y dispara las acciones presionando la superficie con su dedo, de esta forma el jugador experimenta nuevas experiencias, como lo es la interacción háptica.

3.1.6. Reconocimiento por visión. La interacción por visión fue lograda por medio del sistema *CamSpace* [6].

Con la ayuda de esta tecnología, la forma de interacción se da de la siguiente manera: se coloca una cinta en el dedo índice y otra en el pulgar, dichas cintas deben de ser de diferente color cada una, esto con el objetivo de facilitar al sistema la detección de los objetos, el movimiento de los dedos representa al personaje del juego y el contacto entre ellos dispara la acción dentro del juego.

3.1.7. Seguimiento de cabeza. Para lograr el seguimiento de cabeza se utilizó el sistema *WiiDesktop VR* [7] el cual hace uso de un control *Nintendo Wii Remote*, un receptor *bluetooth* y LEDs infrarrojos. El *wii remote* es un control sensible al movimiento equipado con una cámara para infrarrojos. Usándolo como cámara, es posible dar seguimiento a LEDs infrarrojos colocados en la cabeza del jugador. El *wii remote* transmite la información por medio del protocolo *bluetooth* la cual es recibida por medio del receptor *bluetooth* mencionado previamente.

3.2. Prototipo

Para probar las funcionalidades de la arquitectura propuesta fueron creados 2 prototipos. Los cuales se detallan a continuación.

3.2.1. Fallbox. Es un juego 2D (ver figura 3) con perspectiva de primera persona. Su forma de entrada de datos por parte del jugador son el seguimiento de cabeza y el teclado. El objetivo de este juego es que los jugadores muevan su cabeza para controlar al personaje principal del juego en el eje X para lograr esquivar cajas que están cayendo en el escenario.

El juego cuenta con 5 niveles, en los cuales por cada nivel se va incrementando la dificultad y las cajas caen a una velocidad mayor. El jugador cuenta con vida de juego limitada (una barra al 100% al iniciar el juego) la cual se va reduciendo cada que el personaje es golpeado por una caja al caer.

3.2.2. Shooter. Es un juego 2D (ver figura 4) en primera persona. Las características y objetivos del juego son simples, el jugador debe pasar a través de diferentes escenarios en los cuales debe eliminar por disparo a todos los enemigos en un tiempo límite, conforme se avanza en los escenarios el puntaje se va incrementando. De esta forma el jugador llegará al escenario final en donde enfrentará al enemigo principal del juego, al cual deberá derrotar para terminar el juego.



Figura 3. Juego Fallbox.

Las formas de interacción multimodales de *Shooter* son: reconocimiento de gestos, reconocimiento por visión y reconocimiento de pluma.

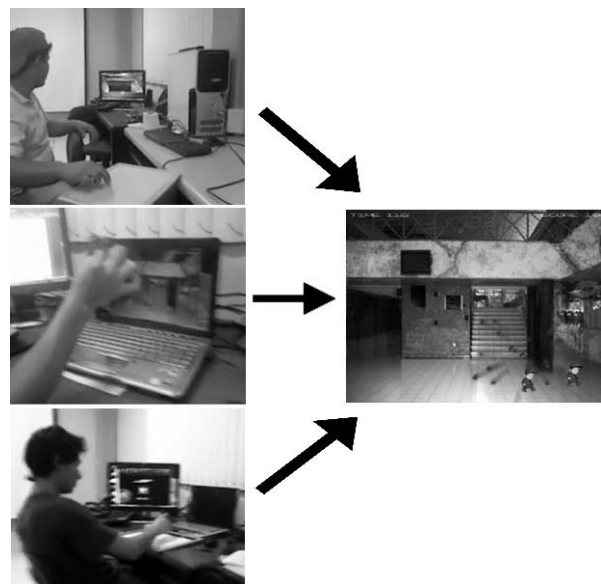


Figura 4. Juego Shooter.

3.3. Evaluación

Para evaluar la arquitectura se realizó una sesión de pruebas con el prototipo *Fallbox* [8], en la cual un número importante de estudiantes y profesores lo probaron; a continuación se describe el proceso de la evaluación y los resultados.

4 Diseño del estudio

El estudio fue realizado en la Facultad de Telemática de la Universidad de Colima. Los sujetos de estudio fueron 20 personas, 11 hombres y 9 mujeres, con un rango de edad entre 14 y 29 años. 70%

de los participantes han tenido experiencia con los juegos por computadora.

Se evaluó la aceptación de uso de la tecnología a través de la experiencia de juego con interfaces multimodales, este prototipo como se explicó anteriormente cuenta con las interfaces de seguimiento de cabeza y teclado.

4.1. Procedimiento

Se realizó una sesión de evaluación de una hora, la cual incluyó las siguientes fases:

Fase 1: Una introducción de 10 minutos.

Fase 2: Se realizó una demostración en vivo para mostrar a los participantes las características del juego.

El objetivo de esta demostración fue poner en contexto a los jugadores con el uso de los dispositivos de interacción.

Fase 3: Se les dio tiempo a los participantes para que libremente usaran la tecnología.

Fase 4: En esta fase, los participantes completaron un cuestionario de experiencia del juego (GEQ por sus siglas en inglés *Game Experience Questionnaire*) con una escala *likert* de 5 puntos, el cual incluyó los temas de eficiencia, efectividad, inmersión, motivación, emoción, fluidez y curva de aprendizaje.

Además se utilizó observación directa en la sesión completa y se tomaron notas sobre los comentarios realizados durante las pruebas.

4.2. Resultados y discusión

Los resultados obtenidos en la evaluación se muestran en la figura 5, los cuales se discuten a continuación.

4.2.1. Obstáculos para la adopción multimodal. Los participantes identificaron una falta de efectividad con el dispositivo de seguimiento de cabeza con respecto al teclado, lo cual la convierte en el principal (y único) obstáculo potencial para la adopción de la tecnología.

4.2.2. Percepción de una experiencia más completa. Los participantes encontraron el uso de la tecnología de seguimiento de cabeza (en relación con el teclado) más eficiente, inmersiva, motivadora, emocionante, fluida y fácil de aprender, lo cual indica que ellos tendrían interés en usar este tipo de dispositivos.

5. Conclusiones

Este trabajo presenta una arquitectura para la construcción de juegos por computadora con interacción multimodal, para lograr una forma más

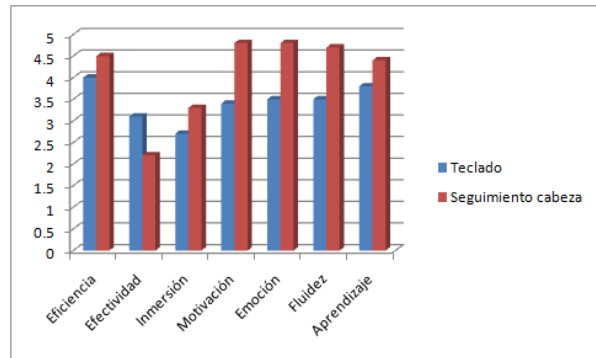


Figura 5. Percepción de los jugadores.

conveniente de obtener los datos E/S para los jugadores.

Los prototipos mostraron el poder y la flexibilidad de esta arquitectura, y permitieron realizar una evaluación de percepción de los jugadores, en la cual se encontró que este tipo de interacciones los involucran más en una experiencia de juego realista.

Con esta arquitectura, se plantea eliminar la limitación principal de las interfaces multimodales, ya que por lo general son diseñadas de forma individual para cada aplicación. De esta manera se cuenta con un modelo que incluye 4 posibles formas de interacción que sólo deben ser instanciadas para cada necesidad en particular del juego a desarrollar, como se demuestra en los 2 prototipos descritos en este artículo.

6. Referencias

- [1] Magerkurth, C., Stenzel, R., Streitz, N., & Neuhold, E. (2003). A Multimodal Interaction Framework for Pervasive Game Applications. Seattle, USA.
- [2] Sharon Oviatt. Ten myths of multimodal interaction. Communications of the ACM. Vol. 42 No. 11, Pages 74-81
- [3] Microsoft. XNA Developer Center. Obtenido de <http://msdn.microsoft.com/en-us/aa937791>
- [4] DUO digital pen. Obtenido de www.penandfree.com.
- [5] Community core vision. Obtenido de <http://ccv.nuigroup.com>.
- [6] CamSpace system. Obtenido de <http://www.camspace.com>
- [7] Lee, J. C. (2008). *Head-tracking for Desktop VR Displays using the Wii Remote*. Retrieved 2010, from <http://www.cs.cmu.edu/~johnny/projects/wii/>
- [8] González, F., Santana, C. P., Calderón, P., Munguía, A., & Arroyo, M. (2010). Fallbox: a computer game with natural interaction through head tracking. *MexIHC 2010*. San Luis Potosí: Universidad Politécnica de San Luis Potosí.