

Arquitectura genérica para la interacción emocional a través de avatares

Amalia Ortiz, David Oyarzun
M^aPuy Carretero, M^aTeresa Linaza, Jorge Posada

Centro de investigación VICOMTech
Paseo Mikeletegi 57 20011 Donostia
aortiz@vicomtech.es

Nestor Garay-Vitoria

Universidad de la UPV/EHU
Manuel Lardizabal 1 E-20018 Donostia
nestor.garay@ehu.es

Resumen

Este trabajo propone la utilización de avatares para generar la comunicación no verbal en un sistema interactivo multimodal, centrándose principalmente en los aspectos emocionales o afectivos de este tipo de comunicación. Para ello, se han analizado todos los aspectos relacionados con la interacción multimodal afectiva a través de avatares, como (i) las principales ventajas y desventajas de integrar un avatar en un sistema interactivo multimodal, (ii) los elementos que componen el sistema interactivo (el usuario, el avatar y la aplicación) y (iii) sus características, (iv) los niveles de interacción que pueden producirse entre los elementos y (v) los tipos de interacción que han de tenerse en cuenta a la hora de generar una interacción multimodal afectiva a través de avatares. Una vez analizada toda esta información se presenta la propuesta de una arquitectura general que engloba todos los tipos de interacción que podrían producirse en un sistema que incluya avatares.

1. Introducción

La interacción persona-dispositivo es un área científica multidisciplinar muy activa, que lleva años trabajando con el fin de hacer que la interacción entre el usuario y los dispositivos sea más natural e intuitiva. Su trabajo es muy importante, debido a que el éxito o fracaso de un sistema interactivo depende, en gran medida, de la interfaz de usuario. Actualmente, los avances en tecnologías como la visión artificial, los gráficos 3D o las tecnologías del habla están

permitiendo una nueva aproximación para la interacción denominada interacción multimodal, que permite un estilo de interacción más natural. Dentro del campo de interacción multimodal, el tipo de comunicación más extendido es verbal. Sin embargo, a la hora de diseñar un sistema interactivo multimodal, hay que tener en cuenta que las personas podemos comunicarnos de dos formas: mediante la comunicación verbal, que nos permite comunicar lo que pensamos mediante palabras, y mediante la comunicación no verbal, que nos permite la comunicación de las emociones.

Además de los lenguajes de la comunicación verbal (lenguaje oral o escrito), que plantean la mayoría de las arquitecturas existentes en el estado del arte, el sistema multimodal debe interpretar y generar los lenguajes relativos a la comunicación no verbal (lenguaje corporal o icónico) para obtener una interacción multimodal natural, de tal manera que el usuario pueda comunicarse con el sistema a través de estos lenguajes y obtener una respuesta de la misma forma. Además, la comunicación no verbal debe estar presente en un sistema interactivo multimodal ya que aporta mucha información al mensaje a transmitir. De hecho, según el estudio de Mehrabian [1], el 93 % del mensaje se transmite a través del lenguaje corporal, básicamente compuesto de expresiones faciales y movimientos corporales (55 %) y el uso de la voz (38 %), pudiendo llegar a cambiar el significado del lenguaje oral.

En la comunicación no verbal, la información se recibe y se interpreta principalmente por dos canales, el emisor, formado por cara, ojos, cuerpo y voz, y el receptor, formado por la vista, el oído, el olor, el sabor y el tacto. Una de las formas de

simular la parte emisora en la comunicación no verbal en un sistema interactivo podría basarse en la utilización de personajes virtuales o avatares. Un avatar es un personaje virtual que está formado por una parte física y otra mental. La parte física puede tener diferentes formas; puede ser 2D o 3D, antropomorfa o tipo *cartoon*, y puede ser una combinación de cara, ojos, cuerpo o voz. La parte mental puede ser inteligente o no. Si el avatar no es inteligente, simplemente visualiza las órdenes que le llegan mediante su cara, sus ojos, su cuerpo o su voz. Si el avatar es inteligente, es capaz de tomar decisiones. A las componentes necesarias para que el avatar pueda tomar decisiones se les suele llamar agentes.

En este trabajo se propone la utilización de avatares para generar en el sistema la comunicación no verbal, centrándose principalmente en los aspectos emocionales o afectivos de este tipo de comunicación. En la sección 2, se analizan las ventajas y desventajas del uso de avatares en la interacción afectiva y en la sección 3 se analizan todos los aspectos relacionados con ella. En la sección 4, se plasma toda esta información en una matriz. A través de esta matriz se realiza en la sección 5 una propuesta para una arquitectura genérica de un sistema que permita interacción multimodal afectiva a través de avatares. Por último, las conclusiones a este trabajo se muestran en la sección 6.

2. Interacción multimodal afectiva a través de avatares

La integración de avatares en diferentes aplicaciones ha sido un campo muy investigado en los últimos años. Diversos autores han realizado evaluaciones con usuarios reales para definir las principales ventajas de la interacción a través de avatares. Entre estas ventajas pueden mencionarse las siguientes:

- Facilita la interacción social con el dispositivo. En 1994, Nass et al. [2] realizaron cinco experimentos que pusieron de manifiesto que las interacciones individuales de las personas con los ordenadores son fundamentalmente sociales. Recientemente, Prendinger et al. en [3] concluyeron adicionalmente que el usuario

espera obtener además el mismo tipo de comportamiento social. Por ello, se propone dotar a la interfaz con aspectos de personalidad y síntesis de voz para mejorar la interacción persona-dispositivo.

- Incrementa la fiabilidad y credibilidad del usuario en el sistema. Para que un usuario tenga la confianza de delegar ciertas tareas en un agente, ha de creer en su fiabilidad. Existen evaluaciones que demuestran que la confianza y la credibilidad aumentan con la personificación del agente, es decir, dándole cara, ojos, cuerpo o voz. Si además el aspecto del personaje es realista, el agente es valorado como más inteligente y agradable [4].
- Incrementa el compromiso del usuario. La personificación de los agentes hace que se incremente el compromiso del usuario con la aplicación. En entornos de aprendizaje, por ejemplo, la personificación del tutor virtual causa un efecto positivo en la percepción de los estudiantes, sobre todo, si además se muestran respuestas afectivas [5].
- Capta la atención del usuario. Hongpaisanwivat et al.[6] concluyeron que el avatar es capaz de captar la atención del usuario y que esta todavía aumenta si el avatar es creíble, ya que se genera la ilusión de vida en el sistema [7].
- Dirige la atención del usuario. Con un avatar se puede dirigir la atención del usuario a los puntos de interés, una ventaja muy valiosa para entornos de educación [3].

Sin embargo, a pesar de estos estudios, existe todavía una gran controversia sobre la idoneidad de los avatares como forma de interacción con el dispositivo, ya que no todas las evaluaciones han obtenido un resultado tan positivo. Por ejemplo, al contrario que en [6],[7],[3] o [8], Walter et al. [9] concluyeron que interactuar con avatares captaba la atención del usuario desde un punto de vista negativo, ya que podía llegar a distraerle.

Sobre la base del estado del arte analizado, el presente trabajo se basa en la utilización de avatares en sistemas interactivos multimodales como elementos transmisores de la comunicación

no verbal a través de su cara, ojos, cuerpo y voz. Aunque las tecnologías relacionadas con el área de personajes virtuales como la animación corporal y la síntesis de voz han sido ampliamente investigados [10] [11], su utilización como un elemento necesario en la interacción multimodal no ha sido prácticamente estudiada. En la mayoría de las aplicaciones que permiten la interacción con avatares, éstos se han integrado como un módulo independiente, que ofrece al usuario información visual y auditiva sobre el sistema. Este aspecto fue estudiado por Not et al. [12] que concluyeron que el avatar no debe ser considerado como una modalidad indivisible, sino como una contribución sinérgica de diferentes canales de comunicación que, correctamente sincronizados, generan una comunicación global. Metafóricamente, su grupo interpreta un avatar como una interfaz multimodal de salida y se estudian todas sus capacidades con el objetivo de generar un lenguaje de marcas que permita una fácil integración de los avatares en cualquier aplicación.

3. Análisis de cada elemento

El objetivo de este trabajo es diseñar una arquitectura genérica que englobe los aspectos de interacción necesarios en un sistema multimodal que integre avatares. Al igual que en el trabajo de Not et al., esta arquitectura también considerará al avatar como una interfaz multimodal, aunque en este caso se plantea una arquitectura con tres elementos que pueden interactuar entre ellos al mismo nivel: el usuario, el avatar y la aplicación. Cada uno de estos elementos podrá existir en el sistema de la siguiente forma:

- Pueden existir uno o varios elementos del mismo tipo, dependiendo del número de usuarios, avatares o aplicaciones en el sistema (ej. entornos colaborativos, ambientes inteligentes multidispositivo,...).
- Cada tipo de elemento posee una interfaz que puede ser multimodal o no. El término multimodal en este trabajo se define como la posibilidad de utilizar dos o más lenguajes de comunicación, ya sea lenguaje escrito, verbal, corporal o icónico.

- Cada uno de los elementos del sistema tiene unas características propias que definen sus capacidades. Estas características permiten clasificar cada elemento como activo o pasivo. Estos conceptos se explicarán con más detalle en posteriores secciones.

Para diseñar la arquitectura general que involucre todos los aspectos de interacción, se considera necesario estudiar primero los elementos que definen en el sistema, sus características, los paralelismos existentes entre cada elemento y los niveles de interacción que podrán producirse en el sistema.

3.1. Características propias de cada elemento

A la hora de diseñar la arquitectura del sistema interactivo multimodal, hay que tener en cuenta la influencia de la forma y funcionalidades de cada uno de los tres elementos. Para diseñar una arquitectura genérica, es importante conocer en detalle las características propias y capacidades de cada elemento. Sin embargo, dado que las características propias del usuario, del avatar y de la aplicación son muy diversas, resulta muy difícil realizar una clasificación completa. Dado que el presente trabajo se centra en la interacción multimodal, se ha realizado una clasificación en función del tipo de comunicación que cada elemento pueda generar o interpretar. De este análisis se detecta que cada elemento puede ser activo o pasivo.

Por un lado, un elemento activo es aquel que puede interactuar directamente con el sistema, bien porque tiene capacidad innata de hacerlo o bien porque el sistema le proporciona las herramientas necesarias para que pueda hacerlo. Por otro lado, el elemento activo es capaz de reconocer los eventos de la interacción, interpretarlos y tomar decisiones. Además, el elemento activo es capaz de hacérselo saber al sistema a través de alguna de sus capacidades de comunicación verbal o no verbal. El elemento pasivo no interactúa directamente con el sistema, sino que delega en el sistema o actúa como intermediario ejecutando ordenes de un elemento activo.

3.2. Niveles de interacción entre cada elemento

La mayor parte de las arquitecturas estudiadas centran su investigación en la interacción entre dos elementos, generalmente el usuario y el avatar [13], [14] o [15]. Sin embargo, en un sistema interactivo multimodal con avatares, el presente trabajo define tres tipos de elementos que interactúan entre ellos y que deberán estar sincronizados: el usuario, el avatar y la aplicación. Para la arquitectura planteada en este trabajo, se definen las siguientes interacciones bidireccionales; Usuario-Usuario, Usuario-Aplicación, Usuario-Avatar, Avatar-Avatar, Avatar-Aplicación, Aplicación-Aplicación. En función de sus características, se establecen dos niveles para cada elemento:

1. Interacción directa. Se establece cuando dos elementos accionan y/o reaccionan directamente. Esta interacción puede realizarse a dos niveles. El nivel 1 se establece cuando uno de los elementos realiza una petición y el elemento con el que interactúa reacciona, pero no es capaz de interactuar con el primer elemento. Por otro lado, el nivel 2 se establece cuando ambos elementos accionan y reaccionan. En este nivel, no existe nada pregrabado, por lo que sólo puede darse entre dos elementos activos.
2. Interacción indirecta. Uno de los elementos actúa como intermediario, es decir, funciona como herramienta de transmisión del mensaje a un tercer elemento con el que se desea interactuar.

3.3. Rol de cada elemento

Los roles de cada elemento se definen en función de sus características y los niveles de interacción entre ellos.

Avatar. El avatar puede desempeñar cuatro roles diferentes dependiendo de sus características y de las características propias del elemento con el que esté interactuando: interactivo (Int), autónomo (Aut), marioneta del usuario (MU) o marioneta del autor (MA). El avatar interactivo es capaz de interactuar con el usuario, con la aplicación y con otros avatares, interpretando cada evento producido

en la interacción y generando la respuesta correspondiente. El avatar puede tomar el rol de interactivo únicamente cuando es un elemento activo y la interacción es directa de nivel 2 o indirecta. Sin embargo, un avatar autónomo es un avatar activo completamente autónomo en la toma de decisiones que interactúa con usuarios o aplicaciones pasivas de manera directa nivel 1. El avatar marioneta del usuario es aquel que ejecuta las órdenes que el usuario le indica no teniendo capacidad de tomar decisiones ya que es un avatar pasivo. Este rol surge cuando la interacción es indirecta o cuando es directa de nivel 1 y el usuario con el que interactúa tiene el rol de autor. Por último, el avatar marioneta del autor es aquel que ejecuta las órdenes definidas por el autor de la aplicación. No tiene capacidad de tomar decisiones, ya que es un avatar pasivo. Este rol aparece cuando la interacción es indirecta o cuando es directa de nivel 1 y el usuario con el que interactúa tiene el rol de espectador.

Usuario. El usuario puede desempeñar tres tipos de roles: interactivo (Int), autor (Aut) o espectador (Esp). Por un lado, el usuario interactivo tomará este rol cuando es capaz de tomar decisiones y el sistema interactivo dispone de las herramientas necesarias para interactuar con el avatar o con la aplicación. El usuario podrá tomar este rol en las interacciones directas de nivel 2. Por otro lado, el usuario autor es el creador del sistema interactivo y quien predefine el comportamiento del avatar y de la aplicación. El autor será un usuario activo que interactúa con el avatar o con la aplicación a través de una interacción directa de nivel 1. Por último, el usuario espectador no interactúa con otros elementos, sino que simplemente observa lo que el creador del sistema ha predefinido o las reacciones de los elementos autónomos.

Aplicación. La aplicación puede jugar tres tipos de papeles: interactiva (Int), autónoma (Aut) o marioneta del autor (MA). El comportamiento de cada rol es similar al del avatar.

4. Tipos de interacción emocional

Los tipos de interacción emocional que existan en el sistema deberán cubrir la necesidad de interpretación y generación de emociones para cada nivel de interacción. Respecto a la interpretación, el usuario posee su propia percepción mientras que el avatar y la aplicación deberán contar con las herramientas informáticas que permitan satisfacer esta interpretación (como reconocedores de voz o gestos). Sin embargo, en la generación de emociones, las características propias de cada elemento (activo o pasivo) van a implicar la utilización de los siguientes tipos de interacción, definidos en este trabajo como:

- Se entiende por interacción emocional estática (IEE) el tipo de interacción necesaria para generar interacción que involucre elementos pasivos. El elemento pasivo recibe y ejecuta acciones pregrabadas; definidas previamente por el creador de la aplicación o grabadas por el elemento activo en el momento de la interacción.
- Se entiende por interacción emocional dinámica (IED) aquella que se genera como resultado del proceso de toma de decisiones del sistema interactivo. En este caso, no existe ninguna emoción pregrabada, sino que todas las respuestas emocionales se generan como resultado de un proceso de evaluación de los eventos de interacción producidos en el sistema.
- Se entiende por interacción emocional humana (IEH) aquella que se genera como resultado del proceso de toma de decisiones del usuario.

Este último tipo de interacción se genera en la propia mente del usuario, quien reconoce, interpreta y evalúa los eventos recibidos, para generar la respuesta emocional. Sin embargo, el sistema necesitará herramientas para poder generar interacción emocional estática y dinámica de los elementos avatar y aplicación. A continuación, se detalla el estado del arte de las herramientas necesarias para de ambos tipos de interacción.

4.1. Interacción emocional estática

La interacción emocional estática se genera como una reacción predefinida en el sistema. El hecho de tener que predefinir el comportamiento emocional implica que el sistema disponga de herramientas para la definición de dicho comportamiento. Esta necesidad ha sido investigada por otros autores que concluyeron que una de las maneras más eficientes de dotar al sistema con cualidades emocionales estáticas es a través de texto etiquetado, es decir, etiquetas que definirán las emociones, intensidades o gestos faciales (en el caso del avatar) que deberá reproducirse en un momento concreto, además del texto que el sistema deberá reproducir.

La necesidad de etiquetar el comportamiento ha llevado a realizar un análisis de los principales lenguajes existentes que fue publicado en [16]. Como puede verse en este análisis, los lenguajes de marcas más completos fueron VHML [17] y RRL [18], ya que incluyen todos los parámetros relacionados con la emoción, la animación facial y corporal, y el etiquetado de diálogos. En este trabajo, se elige por VHML debido a que se trata de un estándar que permite definir todos los aspectos de la interacción emocional estática.

4.2. Interacción emocional dinámica

Para que un sistema interactivo sea capaz de generar emociones, será necesario dotarlo de un módulo capaz de generar emociones automáticamente. Las emociones serán el resultado de un proceso cognitivo evaluador de los eventos en el entorno. Existen varios modelos emocionales que se han basado esta perspectiva cognitiva, como los desarrollados por Aaron Sloman [29], Lazarus [30], Ortony, Clore y Colins [31] o Roseman [32]. Dentro de este conjunto, el de Ortony, Clore y Colins (OCC) y el de Roseman fueron explícitamente pensados para su integración en los ordenadores y ofrecen un mecanismo basado en reglas para la generación de emociones cognitivas. De hecho, muchos autores han desarrollado sistemas computacionales basados en estos modelos emocionales. Por ejemplo, el modelo de Roseman fue implementado por Velasquez en Cathexis [33]. Por su parte, el modelo OCC ha sido extensamente implementado,

Cuadro 1: Matriz de interacción emocional

	Avatar activo	Avatar pasivo	Usuario activo	Usuario pasivo	Aplicación activo	Aplicación pasivo
Avatar activo	» Int « Int Directa N2 » IED « IED [19]	» Int « MU Indirecta » IED « IEE	» Int « Int Directa N2 » IED « IEH	» Atm « Esp No hay X	» Int « Int Directa N2 » IED « IED	» Atm « MA Directa N1 » IED « IEE
Avatar pasivo	» MU « Int Indirecta » IEE « IED [19]	» MU « MU Indirecta » IEE « IEE [20] [19]	» MU « Aut Directa N1 » IEE « IEH	» MA « Esp No Interacción » IEE « IEH(autor)	» MU « Int Indirecta » IEE « IED	» MA « MA Indirecta » IEE « IEE
Usuario activo	» Int « Int Directa N2 » IEH « IED [21]	» Aut/Int « MU/MA Directa N1 » IEH « IEE [16]	X No IPO X	X No IPO X	» Int « Int Directa N2 » IEH « IED	» Aut/Int « MA Directa N1 » IEH « IEE
Usuario pasivo	» Esp « Atm No Interacción X [22]	» Esp « MA No Interacción » IEH(autor) « IEE [23]	X No IPO X	X No IPO X	» Esp « Atm No Interacción X	» Esp « MA No Interacción X
Aplicación activo	» Int « Int Directa N2 » IED « IED [19]	» Int « MU Indirecta » IED « IEE [24]	» Int « Int Directa N2 » IED « IEH [25]	» Atm « Esp No hay X	» Int « Int Directa N2 » IED « IED [26]	» Atm « MA Directa N1 » IED « IEE
Aplicación pasivo	» MA « Atm Directa N1 » IEE « IED [27]	» MA « MA Indirecta » IEE « IEE [28]	» MA « Aut/Int Directa N1 » IEE « IEH [8]	» MA « Esp No interacción X	» MA « Atm Directa N1 » IEE « IED [26]	X X No interacción

destacando los modelos desarrollados por Elliot [34] y por Bates [35]. Otros autores, al encontrar limitaciones en ambos modelos, los han combinado, como El-Nars [36] o Buy [37]. Bartneck apunta en su tesis [38] que el motivo por el que la mayoría de los proyectos en esta área eligen el modelo OCC es que es el único que ofrece una estructura de variables que influyen la intensidad de la emoción.

En una comparativa entre ambos modelos, se han encontrado diferencias significativas que hacen que el uso de cada uno de ellos dependa principalmente del nivel de interacción que se desee obtener en el sistema. Por ejemplo, para las interacciones directas de nivel 2 se ha concluido que el modelo OCC tiene en cuenta los modelos de estándares y preferencias del usuario en su proceso de evaluación, mientras que Roseman sólo evalúa de acuerdo a objetivos y esto hace que algunas emociones relacionadas con actitudes o estándares (gusto/disgusto o enfado) no puedan definirse de una forma concreta. Sin embargo, existen muchos otros casos dentro del conjunto de interacciones indirectas o directas de nivel 1 donde no puede llegar a conocerse el modelo del usuario, por lo que la implementación de Roseman puede ser más adecuada para estos casos.

En general, en este trabajo, se apuesta por el modelo de Roseman por los siguientes motivos. Por un lado, como puede verse en el cuadro 1 existen un mayor número de casos en los que el modelo de Roseman se ajusta mejor. Además, este modelo considera la emoción sorpresa dentro de sus 17 emociones y esta emoción es muy importante ya que está considerada por Ekman [39] como una de las seis emociones básicas universales. Sin embargo, el modelo OCC no la contempla. Otro motivo que nos lleva al modelo de Roseman es que, como Bartneck concluyó, es necesario almacenar una función histórica que ayude a valorar la probabilidad, realización y esfuerzo de cada evento para categorizar las emociones, elemento que no se contempla en el modelo OCC.

4.3. Matriz de interacción emocional

El cuadro 1 presenta y visualiza todos los aspectos analizados en secciones previas (características propias de cada elemento, niveles de interacción, roles de cada elemento y tipos de interacción) para satisfacer las necesidades de la interacción emocional y los visualiza en forma de una matriz simétrica (cuadro 1), incluyendo las interacciones que pueden existir en el sistema y

la forma en la que se producen. Además, pueden observarse algunos ejemplos de aplicaciones publicadas que se encuadran bien en la matriz.

5. Arquitectura genérica

Aunque la utilización de avatares en interfaces de usuario ha sido investigada por varios autores, los niveles de interacción establecidos en arquitecturas existentes tienden a centrarse en la interacción producida entre el usuario y el avatar diseñado para dicha arquitectura.

Un ejemplo es el caso de Adele [13]. En este caso, la interacción entre el avatar y el estudiante se basa en el lenguaje escrito, por lo que la arquitectura de este sistema no considera a la aplicación como un elemento de interacción. Otras arquitecturas que también utilizan avatares son las diseñadas por los autores [14] o [15], que buscan el objetivo de mejorar la interacción usuario-avatar en entornos virtuales mediante la integración de módulos inteligentes que generen un avatar semi-autónomo. En el primer caso, se definen cuatro niveles de interacción: usuario-usuario, usuario-su propio avatar, usuario-avatar de otro y avatar-avatar. La segunda arquitectura plantea dos niveles, usuario-su agente y agente-agente. Dado que este trabajo se centra en imitar la comunicación afectiva entre personas en un entorno virtual, es muy importante centrarse en los aspectos afectivos entre las relaciones que en las emociones de una manera individual. En estos casos, aunque sí contemplan aspectos emocionales, sólo se utilizan en la interacción avatar pasivo-avatar pasivo y tampoco considera la aplicación como un elemento de interacción.

El presente trabajo no pretende diseñar una arquitectura para un tipo de interacción concreta, sino una arquitectura genérica de los tipos de interacción que pueden producirse en un sistema que incluya avatares.

Como ya se ha mencionado previamente, existen tres tipos de elementos en la interacción con avatares. El tipo de comunicación que se establece en el sistema es una combinación de los niveles de interacción producidos entre cada uno de ellos. Para que la integración del avatar en la aplicación sea natural y la respuesta del sistema sea coherente, todas las interfaces que componen el sistema han de

estar correctamente conectadas y sincronizadas.

En la figura 1 se puede observar la arquitectura propuesta. Cada interacción producida en el sistema se realiza directamente entre cada elemento. Las interacciones producidas entre elementos pasivos se representan con líneas discontinuas, mientras que las interacciones producidas por elementos activos se representan por líneas continuas. Los tres elementos involucrados, “habitan” en el sistema con su parte “física” y su parte “mental”. La parte física del usuario será su cuerpo y la del avatar o la de la aplicación será su parte visual. Cada parte física se representa en la arquitectura con el nombre de cada elemento (avatar, usuario o aplicación), seguido de sus características propias (activo o pasivo).

Respecto a la parte “mental”, cada elemento contendrá los módulos necesarios para reconocer e interpretar los datos de entrada, procesarlos y generar la respuesta correspondiente.

En el caso del usuario, la parte “mental” estará compuesta de sus módulos de *interpretación y toma de decisiones*, cuyas capacidades dependen de las características propias del usuario. Los otros elementos del sistema (avatar y aplicación) se comunican con el usuario a través de su parte física (visualización), que éste reconoce a través de la vista (parte física) e interpreta a través de su percepción. Una vez reconocido e interpretado lo que el sistema le quiere comunicar (ya sea a través del avatar o de la aplicación), el usuario toma una decisión y la comunica al sistema a través de su parte “física”.

Para comprender los datos de entrada al sistema, cada elemento posee un módulo *reconocedor e intérprete*, capaz de reconocer e interpretar los eventos de interacción procedentes de la comunicación verbal o no verbal.

Una vez que el sistema ha reconocido e interpretado la petición del usuario, de otro avatar o de otra aplicación, comienza a funcionar la parte “mental”. Tanto el avatar activo como la aplicación activa poseen un módulo inteligente, el *agente* y la *lógica de la aplicación*, que conoce la naturaleza del sistema y sabe cómo reaccionar ante dicha petición. Además, cada elemento activo tiene un módulo asociado capaz de realizar el proceso de toma de decisiones emocionales. En el caso del avatar activo y de la aplicación activa será el módulo compartido *Generador de la respuesta emocional*

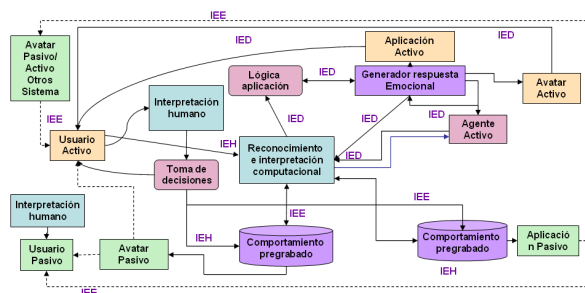


Figura 1: Propuesta de arquitectura general.

quien genere la respuesta emocional coherente al usuario. Por otro lado, respecto a la parte “mental” de los elementos pasivos, tanto el avatar como la aplicación tendrán asociado un módulo contenedor de reglas y eventos emocionales pregrabados. Todas las interacciones producidas entre algún elemento pasivo y algún elemento activo, ya sea a través de interacción emocional estática o interacción emocional dinámica, necesitan reconocedores como única vía capaz de interpretar el lenguaje, ya sea computacional o humano, e indicar a los módulos inteligentes (toma de decisiones, agente o lógica de la aplicación) el significado del evento de interacción producido. También puede observarse en la arquitectura que la interacción entre el agente y la lógica de la aplicación se realiza mediante los módulos de interacción emocional dinámica, aunque incluyendo un módulo generador de la respuesta emocional. La lógica y el agente deciden lo que debe ocurrir en el sistema, se lo comunicarán al generador, que es quien envía el contenido que se debe visualizar a las partes físicas del avatar y del entorno. Cualquier interacción producida en el sistema se representa como un circuito cerrado, ya que todas las interacciones producirán algún tipo de respuesta.

6. Conclusiones

El principal objetivo de este trabajo es la búsqueda de una interacción persona-dispositivo más natural e intuitiva. En el estado del arte, se ha visto que, a pesar de la gran cantidad de contribuciones relacionadas con la comunicación

verbal en los sistemas interactivos multimodales, los aspectos de comunicación no-verbal no han sido tan estudiados. En este trabajo se plantea el uso de avatares como un elemento necesario para obtener una comunicación no-verbal entre el usuario y el dispositivo.

Para estudiar la capacidad de los avatares a la hora de generar comunicación no-verbal, se ha realizado un análisis de los principales aspectos que influyen en la comunicación entre el usuario y el dispositivo cuando intervienen avatares en la interacción multimodal. Como primer resultado de este análisis puede concluirse que existen tres elementos en este tipo de sistemas interactivos cuya interacción ha de estar sincronizada. Estos tres elementos son el usuario, el avatar y la aplicación.

En un análisis más detallado de estos tres elementos, se ha concluido que las características propias de cada uno de ellos condicionan la forma de interacción, de tal manera que dichas características van a definir el nivel de interacción que se producirá entre los elementos y su rol en dicha interacción.

Sin embargo, el principal objetivo del trabajo ha sido estudiar las posibles interacciones a través de la comunicación no-verbal (principalmente aspectos emocionales) que pueden existir en un sistema que incluya avatares. De esta manera, una vez conocidas las características propias de cada elemento, los niveles de interacción y el rol de cada elemento, se han analizado los componentes necesarios que han de existir en un sistema para generar interacción emocional. De este análisis se concluye que son necesarios tres tipos de interacción en el sistema interactivo, IEE (Interacción Emocional Estática),

IED (Interacción Emocional Dinámica) e IEH (Interacción Emocional Humana).

Para mostrar gráficamente toda esta información se ha diseñado una matriz en la que se define, por cada interacción entre dos elementos; el nivel de interacción que puede producirse, el rol que juega cada elemento y los tipos de interacción que han de existir en el sistema para generar interacción emocional, así como un caso de aplicación.

La principal aportación de este trabajo es la creación de una arquitectura genérica diseñada a partir de la matriz de interacción, que incluye todos los módulos necesarios para una interacción multimodal afectiva a través de avatares. La arquitectura ha sido validada por partes en los siguientes trabajos [21], [16] o [23], sin embargo, como trabajo futuro se plantea realizar un sistema modular que permita generar todos los tipos de interacción emocional definidos en la matriz de interacción.

Referencias

- [1] A. Mehrabian. Communication without words. *Psychology Today*, 2(4):53–56, September 1968.
- [2] C. Nass, J. Steuer, and E. R. Tauber. Computers are social actors. In *CHI '94*, pages 72–78, New York, NY, USA, 1994.
- [3] H. Prendinger, C. Ma, J. Yingzi, A. Nakasone, and M. Ishizuka. Understanding the effect of life-like interface agents through users' eye movements. In *ICMI '05*, pages 108–115, New York, NY, USA, 2005.
- [4] T. Koda and P. Maes. Agents with faces: The effects of personification of agents. In *5th IEEE International Workshop on Robot and Human Communication*, Tsukuba, Japan, November 1996.
- [5] Y. Kim. *Pedagogical agents as learning companions: the effects of agent affect and gender on student learning, interest, self-efficacy, and agent persona*. PhD thesis, Tallahassee, FL, USA, 2004. r.
- [6] C. Hongpaisanwiwat and M. Lewis. Attentional effect of animated character. In *INTERACT*, 2003.
- [7] W. S. N. Reilly. *Believable Social and Emotional Agents*. PhD thesis, 1996.
- [8] A. Ortiz, M. Carretero, D. Oyarzun, J. Yaguas, C. Buiza, MF. Gonzalez, and I. Etxeberria. Elderly users in ambient intelligence: Does an avatar improve the interaction? In *User Interfaces For All*, Konigswinter, Germany, 27-28 September 2006.
- [9] R. Subramani J. H. Walker, L. Sproull. Using a human face in an interface. In *Human factors in computing systems*, pages 85–91, New York, NY, USA, 1994.
- [10] T. Giang, R. Moone, C. Peters, and C. O'Sullivan. Real-time character animation techniques. Technical report, Trinity College Dublin, Computer Science Department, 2000.
- [11] A. Aucella, R. Kinkead, A. Wichansky, and C. Shmandt. Voice: technology searching for communication needs. In *CHI '87*, pages 41–44, New York, NY, USA, 1987.
- [12] E. Not, K. Balci, F. Pianesi, and M. Zancanaro. Synthetic characters as multichannel interfaces. In *ICMI '05*, pages 200–207, New York, NY, USA, 2005.
- [13] E. Shaw, W. L. Johnson, and R. Ganeshan. Pedagogical agents on the web. In *AGENTS '99*, pages 283–290, New York, NY, USA, 1999.
- [14] J. Chen X. Mi. Agent-based interaction model for collaborative virtual environments. In *International Conference on Computer Supported Cooperative Work in Design*, pages 401–404. IEEE, May 2005.
- [15] D.L. Ballin M. Gilles. Affective interactions between expressive characters. In *Conference on Systems, Man and Cybernetics.*, pages 1589–1594. IEEE, 2005.
- [16] M.P. Carretero, D. Oyarzun, A. Ortiz, I. Aizpurua, and J. Posada. Virtual characters facial and body animation through the edition and interpretation of mark-up languages. *Computer & Graphics*, 29(2):189–194, 2005.

- [17] A. Marriot. Vhml: Virtual human markup language. In *OZCHI*, Fremantle, Western Australia, November 2001.
- [18] Rrl. <http://www.ai.univie.ac.at>, 2006.
- [19] L. and A. Wachowski. The matrix. Warnerbross film, 31 March 1999.
- [20] SecondLife. Secondlife - your world. your imagination. <http://secondlife.com/>, 2006.
- [21] A. Ortiz, D. Oyarzun, M. Carretero, and N. Garay-Vitoria. Virtual characters as emotional interaction element in the user interfaces. In *AMDO 2006*, Mallorca, Spain, 11-14 July 2006.
- [22] W. Wright. The sims. <http://thesims.ea.com/>, 2004.
- [23] I. Aizpurua, A. Ortiz, D. Oyarzun, I. Arizkuren, J.Posada, A.C. Andrés, and I.Iurgel. Adaptation of mesh morphing techniques for avatars used in web applications. In *AMDO2004*, pages 26–39, Palma de Mallorca, Spain, September 2004.
- [24] M.T. Linaza, A. García, A. Susperregui, and C. Lamsfus. Interactive mobile assistants for added-value cultural contents. In *Proc. of the 7th International Symposium*, 2006.
- [25] C. Malerczyk and M. Schnaider. A mixed reality-supported interactive museum exhibit. In *University College London The Institute of Archaeology*, 2004.
- [26] K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leijten, and J-C. Burgelman. Scenarios for ambient intelligence in 2010. Technical report, European Commission - Community Research, 2001.
- [27] A. Ortiz, D. Oyarzun, I. Aizpurua, I. Arizkuren, and J.Posada. Three-dimensional whole body of virtual character animation for its behavior in a virtual environment using h-anim and inverse kinematics. In *IEEE Computer Society Press, editor, Institute of Electrical and Electronics Engineers (IEEE)*, pages 307–310, 2004.
- [28] MT. Linaza, A. Ortiz, I. Leanizbarrutia, and J. Florez. Collaborative cultural experience- the farm of igartubeiti. In *ARTECH 2004*, Lisboa, Portugal, 12 July 2004.
- [29] A. Sloman. *How many separately evolved emotional beasts live within us.*, chapter Emotions in humans and artifacts. MIT Press, Cambridge, MA., 2002.
- [30] R.S. Lazarus, I.R. Averill, and E.M. Opton. *Toward a Cognitive Theory of Emotion*, chapter Feeling and emotion., pages 207–232. Academic Press, New York, 1970.
- [31] A. Ortony, G.L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, Cambridge, England., 1998.
- [32] I. Roseman, A. Antoniou, and P. Jose. *Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory*. *Cognition and Emotion*, 10, 1996.
- [33] J.D. Velasquez. Modeling emotions and other motivations in synthetic agents. In *AAAI Conference*, pages 10–15, Providence, 1997.
- [34] C. Elliot. *The Affective Reasoner: A process model of emotions in a multi-agent system*. PhD thesis, Northwestern Universtiy, Evanston, IL, 1992.
- [35] J. Bates. The role of emotion in believable agents. *Source Communications of the ACM archive*, 37(Issue 7):122 – 125, July 1994.
- [36] T. Ioerger M. El-Nasr, J. Yen. Fame:fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems*, 3:219 – 257, September 2000.
- [37] D. Buy. *Creating emotions and facial expressions for embodied agents*. PhD thesis, University of Twente, July 2004.
- [38] C. Bartneck. *eMuu - An Embodied Emotional Character for the Ambient Intelligen Home*. PhD thesis, For ambient intelligent home i Technische Universiteit, Oktober 2002.
- [39] P. Ekman. Facial expression and emotion. *American Psychologist*, 48(4):384–392, April 1993.