

EXPRESIONES EMOCIONALES DE AVATERES EN ENTORNOS VIRTUALES COLABORATIVOS

Ildiko Pelczer, Fernando Gamboa y Francisco Caviedes
Laboratorio de Interacción Humano Máquina y Multimedia
Centro de Ciencias Aplicadas y Desarrollo Tecnológico, UNAM
caviedes@aleph.cinstrum.unam.mx

1. INTRODUCCIÓN

Los avances tecnológicos nos permiten actos de comunicación a pesar de las distancias físicas, en tiempo real o de una manera asíncrona. El crecimiento del Internet facilitó el desarrollo de la educación a la distancia, pero en mismo tiempo impuso la necesidad de hacer una revisión de las herramientas tecnológicas a emplear y, al mismo tiempo, generó la necesidad de renovar metodologías pedagógicas. Varios estudios confirman el efecto benéfico de la interacción directa estudiante-tutor o estudiantes-estudiantes sobre el proceso de aprendizaje, ya que facilita una reflexión recíproca sobre tareas y soluciones, motiva a un mejor desempeño y permite el apropiado monitoreo de los alcances del estudiante [1,2].

Los Entornos Virtuales Colaborativos (EVC) representan una solución en este aspecto por su interfaz visual e interactivo. El EVC es un sistema con múltiples usuarios que ofrece una visión inmersiva en sistemas 3D, así como la posibilidad de interactuar en tiempo real con el sistema tal que le permita al usuario generar i) una sensación de espacio compartido ii) una sensación de presencia iii) la sensación de tiempo compartido iv) una manera de comunicarse con sus iguales y v) una manera de compartir [9]. Tal medio ambiente permite fundamentar un aprendizaje socialmente interactivo en lugar de entregar contenidos de manera aislada y sin retroalimentación continua, como pasa en entornos clásicos de educación a la distancia [10].

Para lograr una comunicación exitosa es necesario explorar cuáles son los elementos que la determinan e influyen su calidad. Por ejemplo, hay que tomar en consideración que la comunicación no-verbal juega un papel importante en las interacciones humanas e influye el comportamiento reactivo. Psicólogos afirman que más de 65% de la información intercambiada es por canales no-verbales [3]. La comunicación no-verbal incluye expresiones faciales, mirada, gestos, postura del cuerpo, representación de sí mismo y contacto corporal. Presentamos aquí el resumen de estos elementos, así como se mencionan en [4].

Las expresiones faciales son indicadores de la personalidad y estado emocional, pero también transmiten datos sobre el interés y grado de confianza en la información transmitida (estos aspectos se denominan como significados comunicativos de las expresiones faciales, según [15]). La mirada es una expresión del enfoque de la atención y juega un papel importante en la reglamentación de los turnos en la conversación. La postura comunica actitudes interpersonales y está estrechamente relacionada con el estado emocional los interlocutores. La representación de uno mismo habla de la personalidad que quiere proyectar el usuario. Por contacto corporal se entienden varias formas de contacto físico y no-físico, como por ejemplo, los que inician o terminan una interacción como lo son los gestos de saludo/despida o abrazo, etc.

Es necesario resaltar que la comunicación no-verbal, o por lo menos algunos aspectos de ella, es dependiente del imaginario cultural al que pertenecen los individuos, por ejemplo, el significado del contacto físico es diferente para cada imaginario cultural, en algunas sociedades se saludan con un beso, mientras que en otras este gesto está reservado para personas muy cercanas, y en cada comunidad existen costumbres sociales que reglamentan las expresiones emocionales, (Vg. que tipo de emoción es adecuada expresar en cada situación). Al momento de diseñar un EVC hay que balancear

todos estos aspectos, dado que el EVC representa un espacio de encuentro de individuos que viven a largas distancias geográficas y pertenecen a comunidades con tradiciones y valores diferentes. Este aspecto requiere re analizar las soluciones y los resultados reportados por investigadores de otras culturas antes de adoptar cualquiera de sus propuestas. Cualquiera investigación relacionada con expresiones emocionales en sistemas de cómputo debe considerar el aspecto de dependencia cultural de estas.

En EVC, el usuario normalmente elige una representación humanoide o por lo menos humanizada (avatar) que le permite interactuar en el entorno virtual con las representaciones de los demás usuarios. En [6], los autores mencionan los siguientes enfoques al problema de representar las expresiones no-verbales de sus avatares:

1. controlada de manera directa: la cara y postura del usuario son capturados y el avatar se modifica adecuadamente, en tiempo real.
2. guiada por el usuario: el usuario define los movimientos para efectuar, en este caso, el ratón o el teclado representan las ligas virtuales entre usuario y avatar
3. autónomo: el humano virtual tiene un estado interno que depende de sus metas y de la información sensorial del medio ambiente. El usuario controla este estado interno, mediante la redefinición de las tareas y metas para el avatar.

Independientemente de la solución adoptada, es necesario determinar los elementos, en este caso las expresiones faciales, que permitan la reproducción de un conjunto seleccionado de estados emocionales, por una parte, y la transmisión del significado comunicativo, por otra parte. Esto se refiere a la concordancia de los estados del avatar con las condiciones existentes en el ambiente.

El trabajo de Ekman [14] identificó seis expresiones básicas, correspondientes a las emociones: sorpresa, enojo, miedo, alegría, disgusto y tristeza, así mismo desarrollo un sistema denominada FACS (Facial Action Coding System, [11]) el cual hace la asociación entre la expresión emocional y los grupos de músculos faciales. Aunque su trabajo tiene una extensa aceptación, la implementación en sistemas de tiempo real es compleja, debido al gran número de parámetros que se manejan. En consecuencia, hay un esfuerzo continuo en la comunidad interesada en el desarrollo de avatares, para identificar un número mínimo de elementos que permitan un manejo en tiempo real de la animación facial y que las expresiones emocionales sean fácilmente identificables por los usuarios y/o de otros avatares.

Actualmente existen varios enfoques posibles en tal tarea. Fabri y otros [8] proponen un número reducido de elementos de la cara para lograr diversas expresiones faciales y hacen una evaluación de su propuesta mediante la comparación del porcentaje de emociones reconocidas en su avatar con las emociones presentados por imágenes basados en FACS.

Mientras Fabri propone un conjunto reducido de elementos predefinidos para obtener expresiones faciales, el trabajo de Ruttkay [12] se basa en la exploración de un espacio continuo de expresiones emocionales para identificar regiones con o sin significado emocional, aprender más sobre su estructura y permitir una reproducción fácil de ellas.

Otro enfoque esta representado por el trabajo de Hendrix [13] en cual se intentan identificar los componentes principales de las expresiones para que de tal manera se reduzca el espacio de los parámetros. En su trabajo se hace un análisis de videos con expresiones emocionales producidas de manera voluntaria y se busca un conjunto reducido de elementos faciales necesarios y suficientes para reproducir tales expresiones.

El presente trabajo, es similar en el objetivo a los de Hendrix y Fabri, en el sentido de que se buscan los elementos faciales necesarios para lograr reproducir varias expresiones faciales, por una parte, y por otra parte, para permitir el manejo en tiempo real de estos parámetros identificados.

2. OBJETIVOS.

El primero consiste en verificar la hipótesis de que las expresiones faciales en sistemas de cómputo no pueden ser estáticas por lo tanto: hay que hacer una evaluación del reconocimiento siguiendo la dinámica de la expresión. Se tiene la hipótesis que la expresión animada lleva a un mejor porcentaje de reconocimiento de las emociones. En el trabajo de Fabri [5], las emociones se presentan con la ayuda de fotografías de expresiones emocionales en un solo avatar y con la expresión en su punto de máxima intensidad. Nosotros asumimos que la dinámica de la expresión añade información suplementaria, si no suficiente, si importante para un mejor reconocimiento.

El segundo objetivo del trabajo es el de analizar la influencia de la fisonomía del avatar en el proceso de identificación de la emoción. En el trabajo de Fabri [5] se usa un sólo avatar, sin embargo es necesario saber si los elementos faciales identificados por un personaje bastan para otros avatares. Tal análisis también reporta información sobre la importancia de algunos de los parámetros encontrados.

El tercer objetivo del trabajo es de identificar el número mínimo de componentes faciales que permitan la reproducción de una variedad de expresiones emocionales. Según hemos mencionado en FACS se manejan 46 parámetros de acción facial de bajo nivel, mientras en el Standard MPEG-4 se usan 66. El manejo en tiempo real de todos estos parámetros se vuelve difícil. Por lo tanto, nos proponemos identificar un número pequeño de elementos que permiten reproducir en tiempo real las expresiones emocionales que se puedan utilizar en nuestro trabajo a futuro. Posterior a la identificación de este grupo mínimo de parámetros, se analizará que otras expresiones se pueden lograr con el empleo de estos.

3. MATERIALES Y MÉTODOS

3.1 Agentes y animaciones

Se prepararon 30 secuencias animadas -en formato video para windos (AVI)- de cabezas, correspondientes a 3 avatares diferentes (Fig.1) disponibles en un programa comercial¹, en las cuales se representan para cada uno de ellos (Tabla I) todas las expresiones faciales. Cada secuencia parte de la posición neutra al estado final de la expresión (Fig. 2), y consta de 60 cuadros a razón de 30 cps (2s por animación) de los cuales 30 corresponden al desarrollo de la animación y otros 30 que presentan la imagen final.

Tabla I
Lista de expresiones por Avatar

Alegría
Disgusto
Enojo
Miedo
Molestia
Satisfacción
Sorpresa
Tristeza

The screenshot shows a web application window titled "Eval1". On the left, there is a form with the following fields: "Edad:" with the value "32" in a text box; "Sexo:" with radio buttons for "Femenino" (selected) and "Masculino"; and "Manejo PC:" with checkboxes for "Bueno" (checked) and "Malo". To the right of the form is a large bracketed area containing the text: "Información sobre el Usuario! botón en el momento en el que el avatar tiene la siguiente expresión:". Below the form is a video player showing a 3D avatar of a man with a mustache and a red plaid shirt. The text "Video presentado" is overlaid on the video. To the right of the video player is a button labeled "Inicio" and a larger box containing the text "Botón de Alto" and "Alto!".

Figura 3. Entorno de pruebas

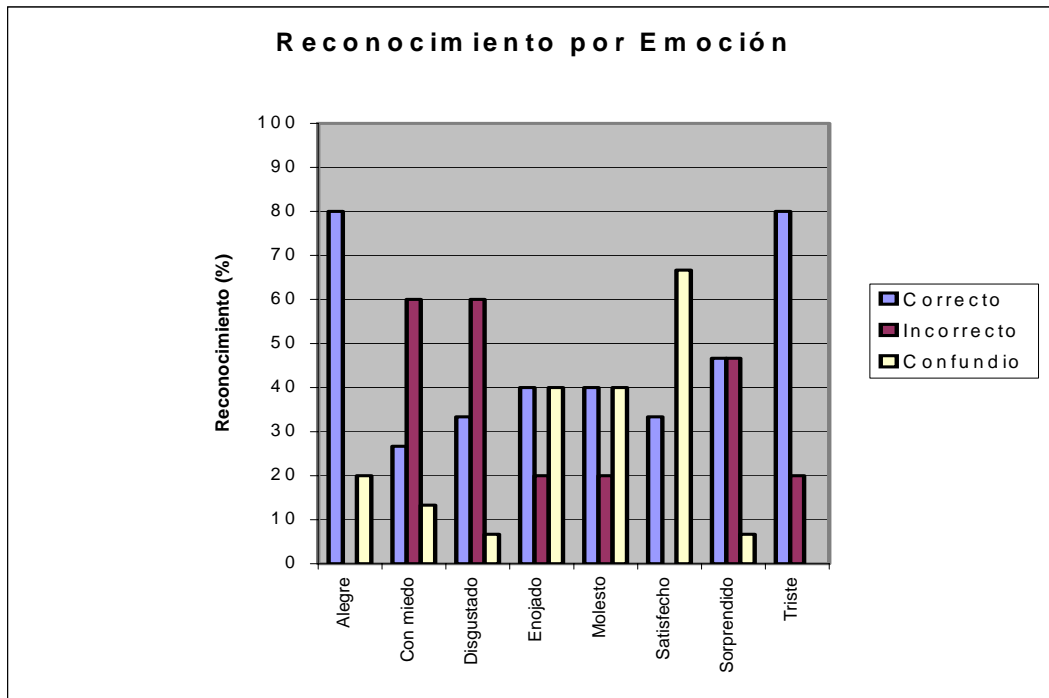
Se presentaron tres animaciones consecutivas y se le solicitó al usuario oprimir el botón de "Alto" en cuanto reconozca una determinada emoción que se le solicita en la ventana. Se le advierte que puede ser la primera, la segunda, la tercera o no aparecer del todo en la prueba, dando la posibilidad de no reconocer ninguna, a pesar de que en cada pregunta se presenta la expresión que se solicita.

Los datos de cada respuesta guardaron la siguiente información en un archivo de texto: Edad, Sexo, Manejo de PC, Expresión solicitada, secuencia señalada y el cuadro en el cual se oprimió el "Alto"

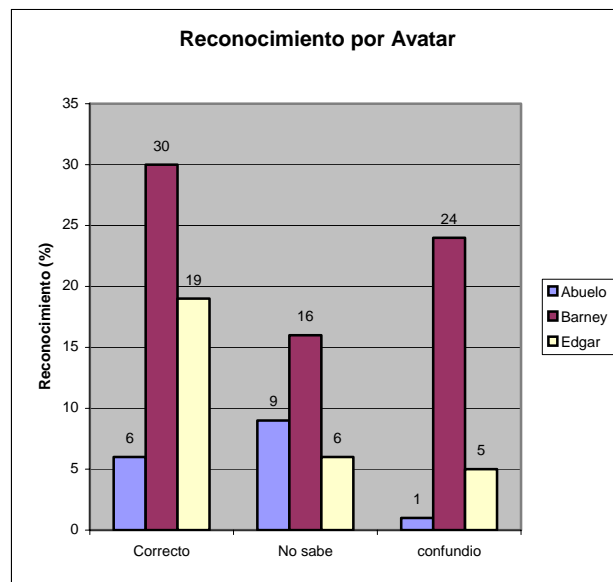
4. RESULTADOS

La muestra consta de 15 individuos con un promedio de edad de 34.4 años consistiendo de 46.6 % femeninos y 53.3% masculinos todos con buena capacitación en el uso de la PC

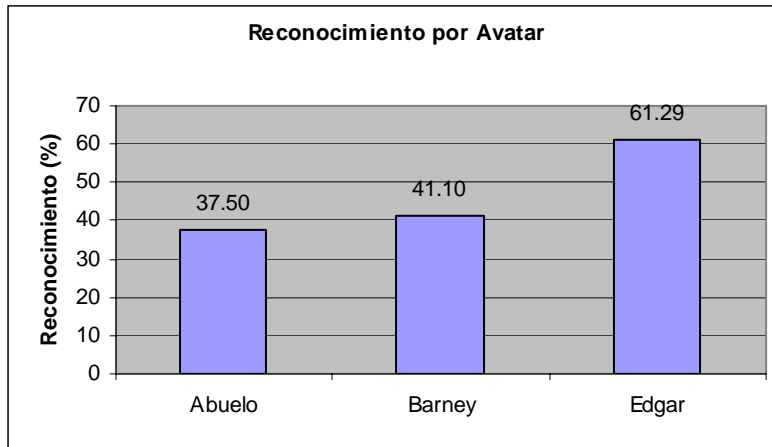
La gráfica 1 muestra el porcentaje de reconocimiento de los usuarios a las expresiones de los avatares.



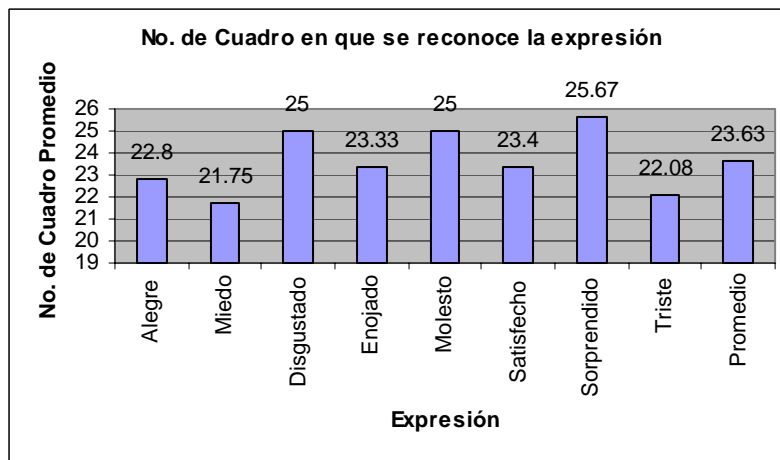
Gráfica 1. Reconocimiento de la expresión desglosada por personaje.



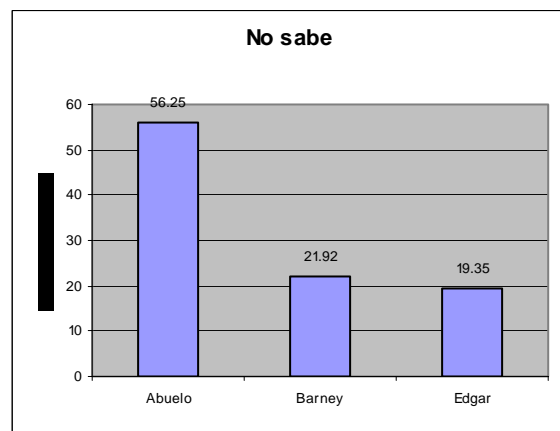
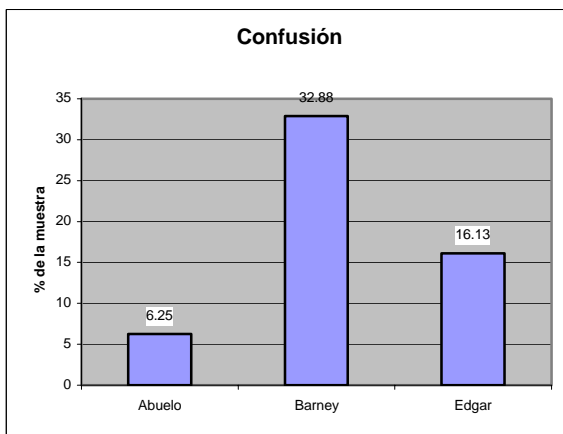
Gráfica 2. Reconocimiento de la expresión desglosada por personaje.



Gráfica 3. Reconocimiento de la expresión, por personaje.



Gráfica 4. Cuadro en cuál se reconoce cada emoción



a) b)
 Gráfica 5. Porcentajes de las expresiones confundidas: (a) y no-identificadas para cada personaje (b)

5. DISCUSIÓN Y CONCLUSIONES

Al analizar los porcentajes de reconocimiento de las expresiones emocionales se encontró un resultado similar con los porcentajes reportados en el trabajo de Fabri et al. Es decir, la hipótesis que la animación ayuda en el reconocimiento de la emoción no se confirmó, sin embargo arroja datos dignos de atención en términos de la velocidad de reconocimiento, los cuales se discuten adelante. Al considerar el valor de la identificación correcta de una emoción en un avatar por parte del usuario (por ejemplo, en sistemas tutoriales) es importante que, para una implementación de un sistema, se busquen factores complementarios tales como, el movimiento de la cabeza o cambio de postura entre otros, lo cual reforzaría la expresión facial y permitiría una identificación correcta del estado emocional del avatar.

El segundo propósito del trabajo era de identificar si hay elementos determinados por la fisonomía del avatar que faciliten o impiden el reconocimiento de la expresión emocional. Las diferencias en el porcentaje de reconocimiento para los diferentes avatares no es significativa (Gráfica 3), sin embargo se constató que en el caso del avatar con el nombre de “Abuelo” el porcentaje de las expresiones no-identificadas es significativamente mayor al porcentaje de los otros dos avatares (Gráfica 5, b).

Al analizar las secuencias animadas con las expresiones no-identificadas para este avatar, se concluyó que la razón puede ser el hecho que la boca de este avatar está ocultado por el bigote, y a pesar de que éste se deforma siguiendo la línea de la boca, interfiere y por lo tanto no se puede seguir adecuadamente la dinámica de los labios y la apertura de la boca. Este resultado sugiere también la importancia de la boca como elemento en la animación facial y se refiere al tercer propósito del trabajo. Por lo tanto, se identifican dos parámetros importantes para la expresión facial: la apertura de la boca y la posición de las cejas.

Un resultado que nos parece particularmente importante, obtenido por el análisis de los tiempos de reconocimiento, es que la expresión emocional se reconoce poco antes que llegue a su expresión máxima o no se reconoce (Gráfica 4). Esto nos lleva a suponer que, si el usuario no intuye hacia donde va la animación; tiende a no reconocerla. Quizás se deba a que, por lo menos en el entorno de la prueba, asumía que el cambio en la expresión tendía a la emoción solicitada causando un prejuicio, en cualquier caso, la velocidad de reconocimiento está por debajo de la duración de la animación. Tal resultado es importante por su implicación en cuanto la duración de la animación de una expresión facial y su desarrollo. Se puede concluir que una expresión emocional que dura menos de un segundo probablemente sea difícil de identificar por el usuario y que si cambia abruptamente, perderá su sentido. Sin embargo cabe señalar que, dado los alcances experimentales del presente trabajo, queda por decidir si la situación contextual en la que se desarrolle la expresión, ayude o perjudice al usuario en el momento en el que se desarrolle.

6. TRABAJO A FUTURO

Las conclusiones también sugieren diversas líneas de trabajo.

En primer lugar habrá que verificar si el motivo del fallo en el reconocimiento no se debe a una expresión facial con pocos elementos. En el proceso de identificación de una emoción a partir de una expresión emocional, los aparentemente usamos pistas provenientes no solamente del movimiento facial sino hay una interpretación del contexto en cuál la expresión ocurre. Habrá que identificar si la dificultad en el reconocimiento se debe al uso de pocos elementos faciales o a la falta del contexto en cuál esto ocurre. Es decir, es necesario usar más componentes faciales para construir las expresiones con bajo porcentaje de identificación y reeefctuar el análisis. Al mismo tiempo, es necesario intentar construir las animaciones usando sólo los factores identificados (cejas, apertura de la boca, etc.) con el fin de verificar si se logra por lo menos el mismo porcentaje de reconocimiento.

Una segunda línea de trabajo debe orientarse hacia la posibilidad de construir con los parámetros identificados otras expresiones emocionales. Así mismo el experimentar con expresiones de emociones

mixtas o co-ocurrentes en el tiempo, las cuales deberían construirse mediante mezclas de expresiones emocionales “básicas” (como las que se analizaron en el presente trabajo).

Hay un tercer elemento esencial para la expresión emocional mediante animaciones: el tiempo de despliegue de la emoción. Es necesario efectuar un análisis detallado con la influencia del tiempo de desarrollo de una expresión emocional y el decaimiento de tal expresión. Por ejemplo, un desarrollo lento de una sonrisa (una animación de larga duración) se puede interpretar como una actitud de burla y no como la expresión de alegría, mientras un tiempo muy corto de decaimiento de la expresión se puede interpretar como una sonrisa falsa.

Al mismo tiempo, es evidente que algunos aspectos complejos de las expresiones emocionales se simplifican una vez que ocurren dentro de un entorno de comunidad co-habitada de avatares/humanos o en implementaciones particulares, y que serán, sin duda, de apoyo al proceso de enseñanza-aprendizaje, tanto en aplicaciones donde los avatares cumplan el papel de tutores, así como en los espacios comunitarios donde los usuarios tengan interacción directa entre ellos.

7. BIBLIOGRAFÍA

1. Webb N.(1995): *Constructive activity and learning in collaborative small groups*. Educational psychology; 87(3); 406-423
2. Wu A, Farrell R., Singley M. Scaffolding (2002): *Group learning in a collaborative Networked Environment*. In: CSCL 2002 Proceedings. Boulder, Colorado.
3. Morris D, Collett P, Marsh P, O'Shauganessy M. (1979): *Gestures, their origin and distribution*. Jonathan Cape ed., London.
4. Katerina Mania, Alan Chalmers. *A classification for user embodiment in collaborative virtual environments*. In: Proceedings of CHI'95 New York, pp. 242–249. ACM Press, (1995).
5. Fabri, M., Moore, D.J., Hobbs, D.J.(2003): *Mediating the expression of emotion in educational collaborative virtual environments: an experimental study*. Virtual reality, vol X(x), Springer Verlag.
6. Fabri, M., Moore, D.J., Hobbs, D.J. (1999): *The emotional avatar: nonverbal communication between inhabitants of collaborative virtual environments*. In: Braffort et al. (eds) *Gesture-based communication in Human-Computer interaction*, Springer Lecture notes in artificial intelligence 1739, pp. 245-248.
7. Fabri, M., Moore, D.J., Hobbs, D.J. (1972): *Face value: Towards emotionally expressive avatars*. In Proceedings of AISB Convention on animated expressive characters for social interaction. London, UK.
8. Fabri, M., Moore, D.J., Hobbs, D.J.(2005): *The use of emotionally expressive avatars* In: CVE. Accepted paper for Symposium on empathic interaction with synthetic characters. AISB 2005, Hertfordshire, April.
9. Sandeep Singhal, Michael Zyda.(1999): *Networked Virtual Environments*. Addison Wesley.
10. Sam Refern, Niall Naughton.(2002): *Collaborative Virtual Environments to support communication and community in internet based distance education* Journal of Information technology education, volume 1, number 3.
11. Ekman, P., Friesen, W.V., Ellsworth, P.(1972): *Emotion in the Human Face*. Pergamon Press, New York.
12. Ruttkay, Zs., Noot, H., Hagen, P.(2003): *Emotion disc and Emotion Squares: tools to explore the facial expression space*. Computer Graphics Forum, 22(1), 2003, pp.49-53.



13. Hendrix, J., Ruttkay, Zs.(2000): *Exploring the space of emotional faces of subjects without acting experience* CWI Report INS-R0013, Amsterdam.
14. Ekman, P., Friesen, W.V.(1978.): *Facial Action Coding System*. Consulting Psychologists Press.
15. Poggi, I., Pelachaud, C. (2000): *Emotional meaning and expression in animated faces*. In: Paiva, A. (ed.) , *Affective Interactions*. Springer, Berlin, NY, pg. 182-195.