

Hacia una representación avanzada del afecto humano para la investigación en calidad de experiencia de usuario

Isabelle Hupont, Eva Cerezo y Sandra Baldassarri

Resumen

Durante los últimos años, la investigación en Computación Afectiva se ha centrado en la extracción automática de las emociones humanas y en incrementar las tasas de acierto en la tarea de reconocimiento emocional en sí. Sin embargo, tras analizar las aplicaciones actuales, se ha detectado que hay una carencia de herramientas automáticas que permitan visualizar de una manera intuitiva la información afectiva extraída del usuario. En el presente trabajo se propone una novedosa herramienta que combina las tecnologías de seguimiento de la mirada y de reconocimiento emocional facial para mostrar representaciones intuitivas y avanzadas de las emociones humanas. El sistema abre las puertas a la realización de nuevos estudios en el campo de la calidad de experiencia de usuario (QoE) que tengan en cuenta aspectos afectivos de los usuarios.

Palabras clave

Computación Afectiva, expresiones faciales, eye tracking, calidad de experiencia de usuario (QoE)

1. Introducción

La Computación Afectiva consiste en el desarrollo de sistemas inteligentes capaces de proporcionar a las computadoras la habilidad de reconocer, interpretar y procesar las emociones humanas [15]. Desde que se acuñó el término Computación Afectiva a finales de los años 90, han aparecido en la literatura numerosos trabajos enfocados a la extracción automática de afecto. Se han desarrollado hasta la fecha sistemas que reconocen las emociones a través de diferentes canales humanos, como expresiones faciales [8], señales fisiológicas [2], voz [7], texto [9], etc. con altas tasas de acierto.

Independiente del canal (o canales) elegido(s) para detectar afecto, la mayoría de trabajos todavía siguen centrando su esfuerzos en incrementar los porcentajes de acierto en la algoritmia de detección emocional. Sin embargo otros factores también muy importantes han sido menos estudiados, por ejemplo cómo visualizar de manera eficiente la información afectiva extraída y cómo procesarla para mejorar la calidad de experiencia (Quality of Experience - QoE) del usuario en diferentes aplicaciones o servicios. La QoE es una medida subjetiva de la experiencia de un consumidor con un servicio o aplicación (navegación web, llamada telefónica, emisión de TV, streaming de video, e-learning...). Investigaciones recientes coinciden en considerar que, dada su naturaleza subjetiva, la medida de factores humanos, en particular de factores afectivos, es una parte esencial para determinar la QoE final percibida por el usuario. Algunos modelos, como el trabajo de Skorin-Kapov y Varela [18], incluso consideran el afecto como una de las dimensiones básicas de la QoE.

En este trabajo se propone una novedosa herramienta para la investigación en QoE afectiva que combina las tecnologías de eye tracking (seguimiento ocular) y reconocimiento de emociones faciales para obtener representaciones gráficas avanzadas del afecto humano. Esta combinación ofrece la posibilidad de relacionar mirada, emociones y contenidos de aplicaciones y servicios, de una manera intuitiva y altamente visual.

La estructura del artículo es la siguiente: la sección 2 analiza el estado del arte relacionado, la sección 3 presenta la herramienta y, finalmente, la sección 4 comprende las conclusiones y trabajo futuro.



2. Trabajo Previo

Esta sección presenta algunas de las investigaciones más recientes en la descripción de emociones y la visualización de la información afectiva, de cara a demostrar el potencial del trabajo propuesto.

2.1 Descripción de las emociones

La forma más usual en que los psicólogos describen las emociones es en términos de categorías discretas. Las categorías más utilizadas son las seis emociones universales propuestas por Ekman [5], que incluyen “alegría”, “tristeza”, “miedo”, “enfado”, “aversión” y “sorpresa”. El esquema de etiquetado basado en categorías es muy intuitivo y concuerda con la forma en que expresamos nuestras emociones en el día a día. Sin embargo, las emociones humanas son mucho más complejas y ricas que simples etiquetas emocionales y pueden experimentar fuertes variaciones a lo largo del tiempo. Estos aspectos de las emociones humanas (complejidad y dinámica) deberían ser capturados y descritos por un reconocedor emocional ideal.

Para resolver los problemas citados previamente, algunos investigadores, como Whissell [21] y Plutchik [16], prefieren ver los estados afectivos no de forma independiente, sino relacionados de una forma sistemática. Ellos consideran las emociones como un espacio 2D continuo cuyas dimensiones son evaluación y activación. La dimensión de evaluación (también denominada valencia) mide cómo se siente una persona, desde muy negativo hasta muy positivo. La dimensión de activación (también denominada excitación) mide si una persona es propensa a realizar una acción bajo un estado emocional, desde muy pasivo hasta muy activo. A diferencia de la aproximación categórica, la aproximación dimensional es atractiva porque provee un álgebra para describir y relacionar un número infinito de estados e intensidades emocionales. Permite trabajar con emociones no discretas y con variaciones de los estados emocionales a lo largo del tiempo. Sin embargo, dada su naturaleza continua (i.e. numérica), la principal desventaja de esta aproximación es que la información afectiva no resulta del todo intuitiva de comprender, dado que las personas solemos reportar nuestras emociones a través de palabras y no de números.

2.2 Visualización de información afectiva

Además del problema de describir las emociones correctamente, una de las carencias principales de los analizadores afectivos actuales está relacionada con el tipo de representación de la información emocional que proporcionan como salida.

Cuando se utiliza una aproximación categórica, la mayoría de los estudios suelen mostrar un histograma o una gráfica representando la distribución -porcentajes o

valores de confianza- de las etiquetas emocionales estudiadas en cada instante de tiempo [8] [6]. Por otra parte, la mayoría de los sistemas que hacen uso de una aproximación dimensional suelen generar gráficas de tipo “activación vs. tiempo” y “evaluación vs. tiempo” [13]. En algunos trabajos, los resultados del análisis emocional están incluso limitados a ficheros de texto simples (logs), que son realmente difíciles de interpretar sin ningún tipo de visualización gráfica [4].

Muy pocos trabajos proponen informes emocionales más sofisticados, visuales y/o intuitivos. Por ejemplo, McDuff et al. [12] presentan el resultado de un analizador de sonrisa a través de gráficas temporales y emoticones, permitiendo comparar el reporte afectivo del usuario con información agregada de otros participantes. El sistema desarrollado en [9] [10] proporciona una representación visual continua de la evolución emocional del usuario dentro del espacio 2D evaluación-activación. Otro ejemplo interesante es el software comercial Affdex© [1], que permite visualizar información emocional dinámica, junto con las expresiones faciales del usuario y los contenidos del vídeo que se ha utilizado como estímulo.

En conclusión, los reportes proporcionados por la mayoría de los reconocedores de emociones se presentan de forma totalmente separada, o como mucho junto al vídeo facial del usuario y los contenidos usados para provocar las emociones. Por lo tanto, no resulta lo suficientemente intuitivo y visual asociar partes específicas, o incluso áreas, de los contenidos a estados emocionales. Sin embargo, existe una necesidad creciente por parte de creadores de contenidos, desarrolladores de aplicaciones, psicólogos o profesionales de Márketing de medir de forma automática la implicación de los usuarios con los contenidos.

La implicación del usuario hace referencia a la calidad de experiencia de usuario, enfatizando los aspectos positivos de la interacción y, en particular, el fenómeno asociado a que el usuario sea cautivado por una aplicación y quiera utilizarla frecuentemente [11]. La implicación es un concepto clave para el diseño de nuevas aplicaciones interactivas, ya que investigaciones recientes demuestran que el éxito de las aplicaciones radica no sólo en que sean utilizadas, sino disfrutadas [14]. Las respuestas emocionales influyen directamente en la implicación de los usuarios [19], y por lo tanto un sistema capaz de reconocer y relacionar de forma eficiente afecto y contenidos resultaría de enorme interés para la investigación en QoE.



3. Una Herramienta Novedosa Para La Visualización Avanzada Del Afecto Humano

Tal como se comentó anteriormente en la sección 2, el afecto y la atención visual son dos marcadores importantes de la implicación del usuario. La atención visual se refiere a la parte (área) de los contenidos que el usuario está mirando en cada momento, mientras que el afecto está relacionado con las emociones experimentadas durante la observación o interacción. Siguiendo esta idea, proponemos una herramienta novedosa y avanzada que permite relacionar de forma dinámica mirada, afecto y contenidos de aplicaciones o servicios. El sistema se basa en la combinación de un eye tracker con un reconocedor de emociones faciales. En la sección 3.1 se presenta la arquitectura y funcionamiento de la herramienta. La sección 3.2 demuestra las amplias y útiles posibilidades de visualización que ofrece para la investigación en QoE.

3.1 Arquitectura y funcionamiento

La herramienta propuesta se construye sobre 2 APIs comerciales que se utilizan ampliamente en nuestro laboratorio de experiencia de usuario desde hace varios años:

Tobii Studio [20] es un software de la compañía Tobii© que ofrece herramientas para crear de una forma sencilla tests y experimentos de eye tracking, obtener datos de la mirada del usuario y construir visualizaciones gráficas a partir de ellos. Tiene un hardware específico asociado, Tobii T60, consistente en un monitor TFT de 17 pulgadas con diodos infrarrojos embebidos que permite la detección en tiempo real de la pupila del usuario. El proceso de eye tracking es no-invasivo para el usuario, admitiendo un rango de movimiento de la cabeza amplio y natural, y cualquier tipo de condiciones de iluminación ambiente. Además, no pierde robustez, precisión, ni fiabilidad en la detección de pupila, independientemente de la raza del usuario, su edad, si usa gafas, etc.

FaceReader [3] es un software de reconocimiento de emociones faciales de Noldus©. Es capaz de analizar en tiempo real las expresiones faciales del usuario, que se capturan a través de una cámara web y, como resultado, ofrece información afectiva tanto a nivel de descripción categórica como dimensional. FaceReader trabaja con gran exactitud y robustez, aún en entornos naturales con cualquier tipo de usuario e iluminación.

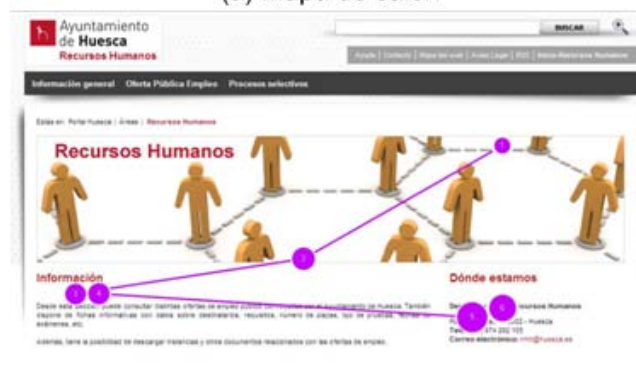
El desafío más importante desde el punto de vista técnico está relacionado con la sincronización de los contenidos analizados, los datos relativos a la mirada y la información emocional. Sin embargo, este problema lo hemos resuelto satisfactoriamente en nuestros primeros test de sincronización.

3.2 Combinando el seguimiento ocular con información de las emociones faciales para investigaciones en QoE

El sistema de seguimiento ocular Tobii© permite construir mapas de calor y rutas sacádicas (Fig. 1). Los primeros son una representación estática del análisis agregado de los patrones de exploración visual de un grupo de usuarios, utilizando un esquema de gradientes de color para mostrar la actividad visual: los colores cálidos revelan las áreas cuyos contenidos han sido más vistos por los usuarios, mientras que los colores fríos muestran aquellas áreas menos vistas por los usuarios. Las rutas sacádicas se utilizan cuando se examina el comportamiento visual de forma individual, indicando el lugar en el cual se centra la mirada el usuario en cada momento, complementado con un pequeño recorrido dinámico que indica los movimientos sacádicos previos.



(a) Mapa de calor.



(b) Mapa sacádico.

Fig. 1. Visualizaciones proporcionadas por Tobii© eye tracker. (a) Ejemplo de mapa de calor. (b) Ejemplo de ruta sacádica.



Por otra parte, el software FaceReader guarda información sobre la actividad afectiva del usuario informando en cada instante de tiempo acerca de la intensidad (porcentaje) de cada una de las seis emociones básicas de Ekman más la “neutral”, y el valor de evaluación (de -1 a 1) que siente la persona. La Fig. 2 muestra un informe típico de FaceReader.

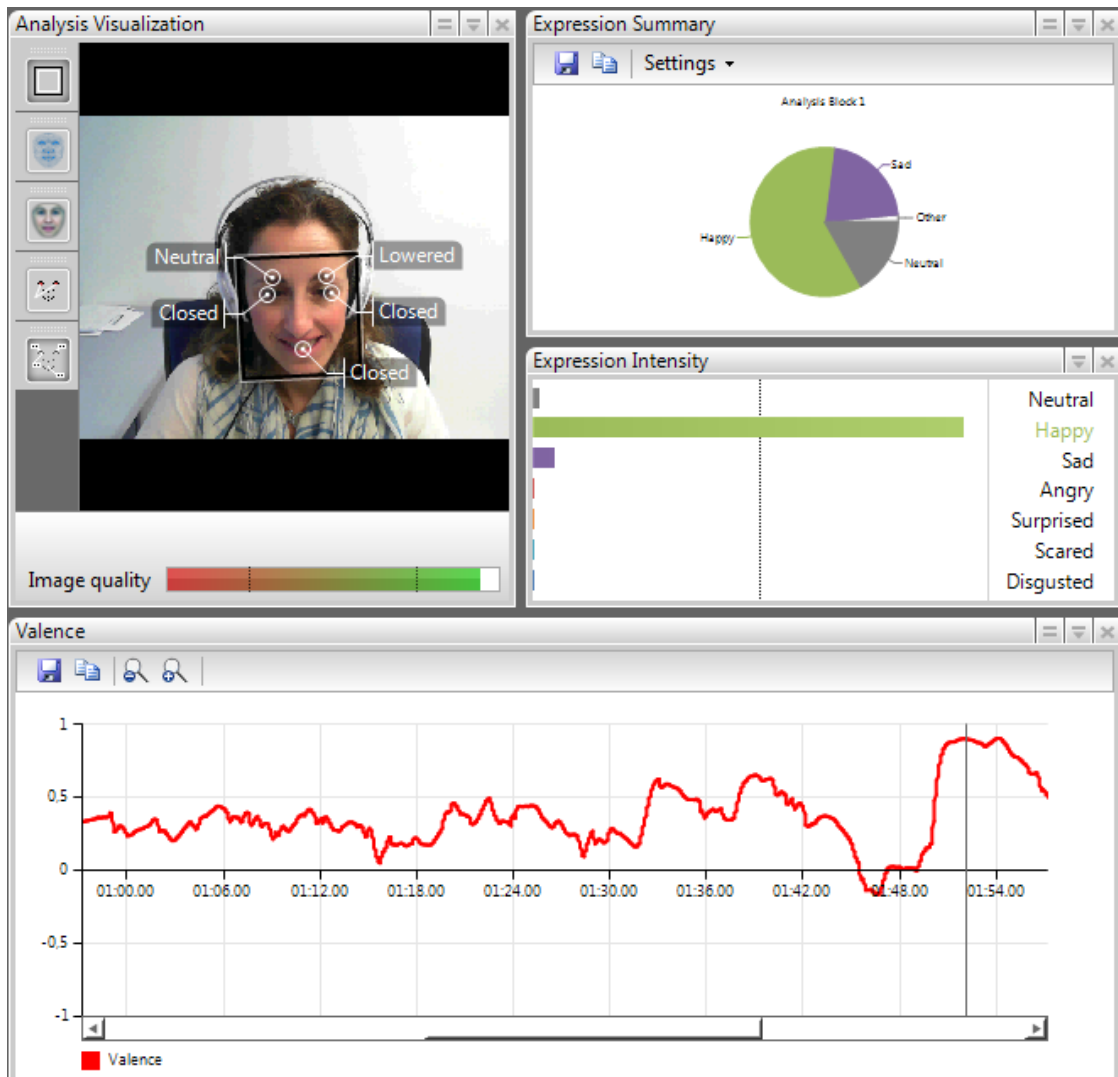


Fig. 2. Típico informe afectivo proporcionado por Noldus© FaceReader.

El objetivo de este trabajo es ir más allá de los mapas proporcionados por el eye tracking tradicional indicando no solo dónde está mirando el usuario, sino también con qué estado emocional. La idea es construir informes visuales en forma de “mapas de calor emocionales” y “rutas sacádicas emocionales” que permitan comprender de un sólo vistazo la relación entre contenidos, mirada y emociones.

Para demostrar el potencial de la herramienta para la visualización del afecto humano, se han llevado a cabo tests preliminares. Se ha construido un videoclip con diferentes imágenes correspondientes a dibujos artísticos. La Fig. 3 muestra

los resultados de una “ruta sacádica emocional” que combina el mapa sacádico del eye tracker con la información emocional extraída por FaceReader para una de estas ilustraciones. En este ejemplo, se sigue el paradigma tradicional de degradado de color del eye tracking para representar los valores continuos de valencia emocional, mientras que las emociones discretas se muestran a través de etiquetas de texto junto a los puntos de fijación visual. De hecho, el principal potencial de la herramienta es el amplio rango de visualizaciones personalizables que ofrece. Su interfaz permite activar y desactivar diferentes opciones de visualización, tanto para los “mapas de calor emocionales” como para las “rutas sacádicas emocionales”, por ejemplo: etiquetas numéricas en los puntos de fijación visual, etiquetas de texto para las emociones discretas, puntos de fijación visual coloreados en función de las emociones discretas, barra de degradado de colores para representar la valencia, etc.



Fig. 3. “Mapa sacádico emocional” construido automáticamente por la herramienta. La información afectiva se visualiza de forma dinámica, tanto en términos discretos como continuos. La barra vertical indica el gradiente de colores correspondiente a la valencia emocional, desde rojo para muy negativa hasta verde para muy positiva.



4. Conclusiones Y Trabajo Futuro

Tras analizar el estado del arte actual en el campo de la Computación Afectiva, se ha resaltado la necesidad de herramientas automáticas capaces de proporcionar representaciones visuales e intuitivas de la información afectiva del usuario. En particular, la investigación en QoE demanda sistemas capaces de relacionar mirada, contenidos audiovisuales y afecto, con la finalidad de obtener retroalimentación sobre el grado de implicación del usuario con las aplicaciones y servicios.

Para satisfacer esta necesidad, el presente artículo propone un sistema novedoso basado en la combinación de las tecnologías de reconocimiento emocional facial y eye tracking. Su principal potencial radica en la posibilidad de obtener visualizaciones avanzadas e intuitivas del afecto humano en forma de “mapas sacádicos emocionales” y “mapas de calor emocionales”. Además, la herramienta ofrece múltiples y altamente customizables opciones para crear representaciones gráficas de mirada, emociones y contenidos en un único mapa.

La herramienta que se propone abre las puertas a la realización de novedosos estudios en el campo de la QoE, y a mejorar aplicaciones y servicios proporcionándoles una capa afectiva. De hecho, este es el trabajo que esperamos realizar en un futuro próximo en el contexto del proyecto europeo QuEEN [17].

Sin embargo, el sistema también resulta de gran interés para muchos otros campos. Por ese motivo, también planeamos explorar su aplicación a la psicología (e.g. para el estudio del autismo, la depresión, la empatía, etc.) y al marketing (e.g. para medir el impacto de campañas de marketing o el análisis de páginas web).

5. Agradecimientos

Este trabajo ha sido parcialmente financiado por la Dirección General de Investigación (TIN2011-24660), por fondos europeos FEDER y por el proyecto Europeo Celtic QuEEN (IPT-2011-1235-430000). Las autoras quieren agradecer a Rakel Goodféith su colaboración artística en el proyecto (<http://goodfeith.blogspot.com>).

6. Referencias

- 1. Affdex, <http://www.affective.com/affdex/> (2013)**
- 2. Chanel, G., Ansari-Asl, K. y Pun, T.:** “Valence-arousal evaluation using physiological signals in an emotion recall paradigm”, Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp. 2662-2667 (2007)
- 3. Den Uyl, M.J. y Van Kuilenburg, H.:** “The FaceReader: Online facial expression recognition”, Proceedings of Measuring Behavior 2005, Wageningen, The Netherlands, pp. 589-590 (2008)
- 4. D’Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., El Kaliouby, R., Picard, R.W. y Graesser, A.:** “AutoTutor Detects and Responds to Learners Affective and Cognitive States”, Proceedings of the Workshop on Emotional and Cognitive Issues at the International Conference of Intelligent Tutoring Systems, pp. 23-27 (2008)
- 5. Ekman, P., Dalglish, T. y Power, M.:** Handbook of Cognition and Emotion (1999)
- 6. eMotion©: emotion recognition software, <http://www.visual-recognition.nl/Demo.html> (2013)**
- 7. Graciarena, M., Shriberg, E., Stolcke, A., Enos, F., Hirschberg, J. y Kajarekar, S.:** “Combining prosodic lexical and cepstral systems for deceptive speech detection”, Proceedings of the International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. 1033-1036 (2006)
- 8. Hammal, Z., Couvreur, L., Caplier, A. y Rombaut, M.:** “Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model”, International Journal of Approximate Reasoning, vol. 46, no. 3, pp. 542-567 (2007)
- 9. Hupont, I., Ballano, S., Cerezo, E. y Baldassarri, S.:** “From a discrete perspective of emotions to continuous, dynamic and multimodal affect sensing”, Advances in Emotion Recognition, ISBN 978-1118130667, Wiley-Blackwell (2013)



- 10. Mandryk, R.L. y Atkins, M.S.:** “A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies”, *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 329-347 (2007)
- 11. McCay-Peet, L., Lalmas, M. y Navalpakkam, V.:** “On saliency, affect and focused attention”, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, pp. 541-550 (2012)
- 12. McDuff, D., El Kaliouby, R. y Picard, R.:** “Crowdsourcing facial responses to online videos”, *IEEE Transactions on Affective Computing*, vol. 3, no. 4, pp. 456-468 (2012)
- 13. Nicolaou, M.A., Gunes, H. y Pantic, M.:** “Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space”, *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 92-105 (2011)
- 14. Overbeeke, K., Djajadiningrat, T., Hummels, C., Wensveen, S. y Frens, J.:** “Let’s make things engaging”, *Funology*, Kluwer (2003)
- 15. Picard, R.W.:** *Affective Computing*, The MIT Press (1997)
- 16. Plutchik, R.:** *Emotion: A psychoevolutionary synthesis*, Harper & Row (1980)
- 17. QuEEN European Celtic-Plus project,** <http://www.celticplus.eu/Projects/Celtic-projects/Call8/QUEEN/queen-default.asp> (2013)
- 18. Skorin-Kapov, L. y Varela, M.:** “A multi-dimensional view of QoE: The ARCU model”, *Proceedings of MIPRO 2012*, Opatija, Croatia (2012)
- 19. Teixeira, T., Wedel, M. y Pieters, R.:** “Emotion-Induced Engagement in Internet Video Ads”, *J. Marketing Research*, vol. 49, no. 2, pp. 144-159 (2010)
- 20. Tobii T60,** <http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-t60t120-eye-tracker/> (2013)
- 21. Whissell, C.M.:** “The Dictionary of Affect in Language”, *Emotion: Theory, Research and Experience*, vol. 4, Academic (1989)

Sobre los autores

Isabelle Hupont obtuvo el título de Ingeniera de Telecomunicaciones en 2006 y de Doctora en Ingeniería Informática en 2010 por la Universidad de Zaragoza. Desde 2007 trabaja como investigadora en la división de tecnologías multimedia del Instituto Tecnológico de Aragón, donde lidera la línea de I+D sobre interfaces multimodales persona-máquina. Es también profesora asociada de ingeniería informática en la Universidad San Jorge de Zaragoza. Sus áreas de interés incluyen la interacción persona-máquina, la computación afectiva, la visión por computador, la inteligencia artificial y los mundos virtuales interactivos. *ihupont@ita.es*

Eva Cerezo es licenciada en Ciencias Físicas y Doctora en Ingeniería Informática por la Universidad de Zaragoza. En la actualidad es profesora titular en el Departamento de Informática e Ingeniería de Sistemas de dicha universidad y fundadora del grupo de investigación GIGA AffectiveLab. Dentro del mundo de la interacción persona-ordenador son la computación afectiva, la interacción multimodal, la interacción tangible y los humanos virtuales. Es autora de más de 70 publicaciones internacionales que acumulan en conjunto más de 300 citas. Es editora de la revista “Advances in Human-Computer Interaction” y, miembro del comité de programa o revisora habitual de numerosas conferencias nacionales e internacionales como el Congreso Internacional de Interacción Persona Ordenador de la AIPO, el “ACM CHI Conference on Human Factors in Computing Systems” y la “International Conference on Intelligent User Interfaces”. *ecerezo@unizar.es*

Sandra Baldassarri es licenciada en Ciencias de la Computación por la Universidad de Buenos Aires (Argentina) y doctora en Ingeniería Informática por la Universidad de Zaragoza (España) en 2004. Desde 1996 es profesora en el Área de Lenguajes y Sistemas Informáticos de la Universidad de Zaragoza (España), actualmente como Contratada Doctora. Es miembro fundadora del grupo de investigación GIGA AffectiveLab y miembro del Instituto de Investigación en Ingeniería de Aragón (I3A), ambos de la Universidad de Zaragoza. Sus áreas de investigación incluyen la computación afectiva, las interfaces multimodales, las interfaces naturales y tangibles, los personajes virtuales emocionales, las diferentes técnicas de interacción aplicadas a la educación y, en particular a la educación especial. En estas áreas de trabajo ha publicado numerosos artículos en revistas y congresos, tanto nacionales como internacionales. *sandra@unizar.es*





© Leonardo Infante

Hacia una representación avanzada del afecto humano para la investigación en calidad de experiencia de usuario

Durante los últimos años, la investigación en Computación Afectiva se ha centrado en la extracción automática de las emociones humanas y en incrementar las tasas de acierto en la tarea de reconocimiento emocional en sí. Sin embargo, tras analizar las aplicaciones actuales, se ha detectado que hay una carencia de herramientas automáticas que permitan visualizar de una manera intuitiva la información afectiva extraída del usuario. En el presente trabajo se propone una novedosa herramienta que combina las tecnologías de seguimiento de la mirada y de reconocimiento emocional facial para mostrar representaciones intuitivas y avanzadas de las emociones humanas. El sistema abre las puertas a la realización de nuevos estudios en el campo de la calidad de experiencia de usuario (QoE) que tengan en cuenta aspectos afectivos de los usuarios.