Universidad de Castilla-La Mancha

Departamento de Sistemas Informáticos

Mecanismos de Interacción Enriquecidos con Técnicas de Computación Afectiva

TESIS DOCTORAL

Presentada por:
José María García García

Dirigida por:
Víctor Manuel Ruiz Penichet (UCLM)
María Dolores Lozano Pérez (UCLM)

Diciembre, 2022

Universidad de Castilla-La Mancha

Departamento de Sistemas Informáticos

Mecanismos de Interacción Enriquecidos con Técnicas de Computación Afectiva

TESIS DOCTORAL

Presentada por:

José María García García

Dirigida por:

Víctor Manuel Ruiz Penichet (UCLM)

María Dolores Lozano Pérez (UCLM)

Diciembre, 2022

Quisiera dedicar este trabajo a mi familia y amigos, que me han acompañado en lo que decidíamos si yo acababa la tesis o la tesis acababa conmigo.

Declaración

Este trabajo es el resultado de mi propio trabajo y no incluye ningún resultado de trabajos hechos en colaboración, excepto donde ha sido específicamente indicado en el texto. Este trabajo no ha sido previamente enviado, ni parcial ni totalmente, a ninguna universidad o institución de ningún grado u otra titulación. Además, por la presente declaro ser uno de los principales autores de los trabajos utilizados en esta tesis por compendio de publicaciones, incluyendo los siguientes trabajos que han sido publicados en revistas con índice de impacto:

- Jose Maria Garcia-Garcia, Victor M. Ruiz Penichet, María Dolores Lozano, Juan Enrique Garrido, Effie Lai-Chong Law, "**Multimodal affective computing to enhance the user experience of educational software applications**", Mobile Information Systems, doi: 10.1155/2018/8751426, 2018 (IF: 1.802, Q3 en 2018)

- Jose Maria Garcia-Garcia, Victor M. R. Penichet, Maria D. Lozano, Anil Fernando, "**Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions**", Universal Access in the Information Society, doi: 10.1007/s10209-021-00818-y, 2021 (IF: 2.629, Q3 en 2021)

- Jose Maria Garcia-Garcia , Maria Dolores Lozano, Victor M. R. Penichet, Effie Lai-Chong Law, "**Building a three-level multimodal emotion recognition framework**", Multimedia Tools and Applications, doi:10.1007/s11042-022-13254-8, 2022, (IF: 2.577, Q2 en 2021)

José María García García
Diciembre 2022

Agradecimientos

En primer lugar, me gustaría agradecer a mis directores, Víctor y María, su constante apoyo, motivación y ánimo. De no ser por ellos, en algunos momentos me habría rendido. Gracias a los dos por estar ahí, por ayudarme a mantener la cabeza firme y centrada en mis objetivos y por animarme a aprovechar la oportunidad que suponía el doctorado cuando terminé el Grado.

Quiero agradecer también a todos los miembros del Grupo ISE (Interactive Systems Engineering) y al director de eventos sociales del mismo (yo) por esos momentos tan buenos que permitían convertir un día de la semana cualquiera en un pequeño acontecimiento que esperar con ilusión.

Finalmente, quiero dar las gracias a las instituciones que me han permitido desarrollar este trabajo: a la Universidad de Castilla-La Mancha y a su Plan Propio de Investigación, que me dio acceso a un contrato que me ha permitido trabajar en mi tesis y en mi futuro de forma exclusiva; al Ministerio de Ciencia, Innovación y Universidades, que ha financiado parte de mi actividad investigadora; y al Banco Santander, que me ofreció una beca para trabajar codo con codo con investigadores de la Universidad de Cambridge durante una estancia internacional.

Resumen

Abstract

Affective Computing (AC) has gone through an exponential growth since it was first proposed in a technical report at the M.I.T by Rosalind Picard in 1995. A lot has changed since that first paper proposed studying human emotions using the power of computers. Years have passed since that moment and, today, Affective Computing is a full field on its own, bringing together disciplines like Machine Learning, Psychology, Medicine, Signal Processing, Human-Computer Interaction, etc.

Practical applications of Affective Computing are as multidisciplinary as its foundations, being possible to apply them on fields such as Marketing, E-learning, E-health, etc. Improving customer service, studying buyers' tendencies according to their mood, analysing students' emotions during lessons, even allowing people to study their own emotions for the sake of their own well-being. This is just a small sample of what researchers are studying right now in the field of Affective Computing, but these applications do not end there.

Over the course of this thesis, which belongs on the field of Human-Computer Interaction (HCI), we have studied AC, elements, tools and software which fall under this discipline, how can they be applied in different use cases and how can we contribute to such a discipline. Due to our previous expertise in this field, we have paid special attention to the fields of E-learning and E-health.

Regarding the contribution of this work, this thesis by compendium has explored different types of contribution to the field of AC as a whole:

- **Application of AC mechanisms to enrich serious games**. For the first publication presented in this thesis, we explored the relationship between emotion and cognition, and the role emotions play in the learning process. By combining emotion detection techniques with a serious game, we managed to assemble a game to teach English to Spanish children which is capable of modifying its difficulty dynamically according to the affective state of the users, finding a balance between boredom and frustration to keep children engaged in the learning process for longer. This proposal was validated

in collaboration with a local school, where a between-subjects test helped validate the initial hypothesis.

- **Application of AC mechanisms to ASD therapies**. Thanks to a collaboration with a local association which works with children with Autism Spectrum Disorder, we were able to study how emotion detection could be used to assist therapists working with these children. One of the obstacles people with ASD have to face is the inability or difficulty to both identify their own emotions and the emotions expressed by other people. However, this skill can be learned with the proper therapy. In order to use AC technologies to assist in this task, we designed an application where children can train their emotion recognition and expression skills in an independent way, but assisted by a therapist if necessary. We validated this proposal with the aforementioned association and received a very positive feedback.

- **Proposal of an architecture to develop multimodal detectors**. Although they are recognised as superior, the design of multimodal detectors involves several difficulties: communication with different services, handling different response formats, merging heterogeneous data, etc. As per our third proposal, we have designed an architecture to organise multiple emotion detectors working together. This proposal includes steps to translate different results to the same format and to merge data using different strategies. This architecture, which has been developed exploiting the features of JavaScript and Node, has been designed in such a way so it can be modified and extended easily, using JavaScript modules.

Finally, even though it is not part of this compendium since it has not been published yet, we decided to take a step further in the development of affective systems and we proposed an extension of the quality model for software proposed in the SQuaRE norms, i.e., the one proposed in ISO 25010. This ISO proposes a set of characteristics and sub-characteristics to measure the quality of software. Since quality models are designed with a high level of abstraction in mind, and due to the novelty of AC, existent models can be hard to customize and adapt to fit these new applications and technologies which did not exist when the model was proposed. To overcome this barrier, we propose an extension of this model based on sub-characteristics and metrics which fit the existent model, so people working the field of AC can measure the quality of affective software highlighting the affective capabilities of it.

In conclusion, we have made some proposals regarding how to create new interaction mechanisms which integrate AC technologies and HCI techniques together. The goal of this integration is to create new ways for users to interact with the applications they use on their daily activities, granting these applications, for instance, the capability of modifying their behaviour based on users' emotions. This thesis also strives to bring attention to the actual application of AC, in contrast with current trends which tend to focus on improving the accuracy of emotion recognition models without actually applying them.

As per our next steps, we are starting to study how to include emotion detection during physical rehabilitation processes, while we study the relationship between emotions, motivation and rehabilitation times. Although this line of work is relatively new, we have already published part of this new work and we are working to extend it.

It is worth noting that this work has been financially supported by several entities. On the one hand, it has been supported by the University of Castilla-La Mancha, through a predoctoral grant published in the Spanish DOCM with announcement number 2018/12504 and resolution number 2019/451. On the second hand, this thesis has been supported by the national project granted by the Spanish Ministry of Science, Innovation and Universities with reference RTI2018–099942-B-I00 and by the regional project (ref: SBPLY/17/180501/000495) granted by the regional government (JCCM) and the European Regional Development Funds (FEDER).

Resumen

La Computación Afectiva (CA) ha crecido de manera exponencial desde que fue mencionada por primera vez en 1995 en un artículo del M.I.T escrito por Rosalind Picard. Muchas cosas han cambiado desde aquel primer artículo que proponía estudiar las emociones humanas usando el poder de la Informática. Años más tarde, la Computación Afectiva es un campo de pleno derecho que integra Aprendizaje Automático, Psicología, Medicina, Detección de señales, Interacción Persona-Ordenador, etc.

Las aplicaciones de este campo son tan multidisplinares como sus fundamentos, teniendo cabida en Marketing, en Educación, en Salud, etc. Mejorar la atención al cliente, estudiar tendencias de compradores en base a su humor, analizar el estado de ánimo de un grupo de estudiantes, de una persona en rehabilitación, permitir a una persona analizar sus propias emociones: esto es solo una muestra de lo que se está haciendo con la Computación Afectiva en estos momentos, pero sus aplicaciones no acaban ahí.

Durante el desarrollo de esta tesis, se han estudiado los distintos elementos afectivos que pueden utilizarse a nivel tecnológico (sensores de señales fisiológicas, detectores de emociones, avatares virtuales, clasificadores automáticos) y diversas formas de aplicarlos. Dada la experiencia previa de nuestro grupo de investigación, estas aplicaciones se han enfocado particularmente en el campo de la salud y la educación.

En lo que respecta a las aportaciones de este trabajo al campo de la CA, en este compendio se han desarrollado distintos tipos de contribuciones:

- **Aplicación de mecanismos de CA para el enriquecimiento de juegos serios**. Para la primera publicación presentada en esta tesis se decidió explorar la relación entre emoción y cognición y el papel que juegan las emociones en el proceso cognitivo que es aprender. Al integrar técnicas de detección de emociones en un juego serio, diseñado para enseñar inglés a niños españoles, se pudo desarrollar un juego capaz de modificar su dificultad dinámicamente en base al estado afectivo de los usuarios. El objetivo de este juego es encontrar, de forma automática, un equilibrio entre el aburrimiento y la frustración para mantener a los usuarios concentrados en el juego durante más tiempo, maximizando así las actividades didácticas que completen. Esta propuesta fue validada en colaboración con un colegio local, donde una prueba usando un grupo de control y un grupo experimental nos permitió validar la hipótesis inicial.

- **Aplicación de mecanismos de CA en terapias con niños en el espectro autista**. Gracias a una colaboración con una asociación local que trabaja con niños con Trastorno del Espectro Autista (TEA) pudimos estudiar de primera mano cómo podían usarse las emociones para asistir a terapeutas en sus actividades educativas con dichos niños.

Uno de los obstáculos que enfrentan las personas con TEA es la dificultad (o incapacidad) para identificar sus emociones y las de otros. Sin embargo, esta habilidad, como cualquier otra, puede desarrollarse con la terapia adecuada, especialmente si se comienza en la infancia. Para poder asistir a personas y terapeutas en sus procesos terapéuticos usando tecnologías de CA se diseñó una aplicación para niños en la cual los usuarios pueden desarrollar sus capacidades para identificar y expresar emociones de forma autónoma, sin perjuicio de que un terapeuta puede guiarlos en el proceso. Esta propuesta se validó junto con la asociación colaboradora mencionada anteriormente, que devolvió comentarios muy positivos.

- **Propuesta de una arquitectura para desarrollar detectores multimodales**. Si bien los detectores de emociones multimodales son considerados como los detectores más completos y fiables, su diseño e implementación conlleva algunas dificultados: comunicar la aplicación con distintos servicios de detección a la vez, sincronizar las respuestas de estos, adaptar el formato de cada respuesta a un formato común, etc. Como tercera propuesta, se ha diseñado una arquitectura para organizar distintos tipos de detectores de emociones operando a la vez. Esta propuesta incluye un flujo de trabajo en el que se traducen los resultados afectivos a un formato común y se integran usando distintas estrategias de fusión de datos. Esta arquitectura, que se ha implementado aprovechando las características de JavaScript y Node.js, se diseñó de forma que fuera fácil de extender y modificar, de manera que pudiera seguir siendo adaptada en el futuro por desarrolladores ajenos a su diseño.

Por último, aunque esta publicación no forma parte del compendio dado que aún no ha sido publicada, decidimos dar un paso más allá en la asistencia al desarrollo de sistemas con tecnología afectiva, por lo que se propuso una extensión del modelo usado para medir la calidad del software propuesto en la ISO 25010, dentro del conjunto de normas SQuaRE. Esta ISO propone una serie de características y subcaracterísticas a través de las cuales se puede medir la calidad del software. Si bien los modelos de calidad existentes presentan un alto nivel de abstracción para ser capaces de adaptarse a todo tipo de software, esta abstracción puede acabar siendo perjudicial, hasta el punto de que se pierdan matices y aspectos importantes del software que sea interesante conocer. Para intentar corregir esta deficiencia, se han propuesto una serie de subcaracterísticas y métricas que, si bien permiten evaluar esos aspectos novedosos de los prototipos con tecnologías afectivas, al mismo tiempo se integran perfectamente con el modelo mencionado anteriormente, lo que avala la coherencia de los nuevos elementos propuestos. Esto permitirá a desarrolladores e investigadores estudiar la calidad de aplicaciones afectivas sin que el modelo borre las capacidades afectivas de las aplicaciones y su influencia en la calidad.

En definitiva, en esta tesis se han realizado diversas propuestas con el objetivo de implementar nuevos mecanismos de interacción que integren la tecnología afectiva con técnicas de Interacción Persona-Ordenador, con la finalidad de modificar la forma en la que las personas interactúan con los dispositivos que forman parte de su día a día, dotando a estos de capacidades para ajustarse y adaptarse a sus usuarios y, más concretamente, a su estado afectivo. Este trabajo persigue también la aplicación práctica de técnicas afectivas, puesto que la aplicación real tiende a quedar como trabajo pendiente en muchos trabajos existentes en esta rama de investigación.

En cuestión de trabajos futuros, se están analizando diversas formas de incluir detección de emociones en procesos de rehabilitación física, como parte del estudio de la relación entre emoción, motivación y tiempo de recuperación. Parte de este trabajo ya ha sido publicado en congresos internacionales y se está preparando para presentarse en otros medios.

Por último, es importante mencionar que esta tesis ha sido financiada por diversas entidades a lo largo del tiempo. En primer lugar, por la Universidad de Castilla-La Mancha, a través de un contrato predoctoral para la formación de personal investigador en el marco del Plan Propio de I+D+i, cofinanciado por el Fondo Social Europeo, publicado en el DOCM con número de anuncio 2018/12504 y número de resolución 2019/451. En segundo lugar, a través de un proyecto nacional concedido por el Ministerio de Ciencia, Innovación y Universidades con referencia RTI2018–099942-B-I00 y de un proyecto regional con referencia SBPLY/17/180501/000495 concedido por la Junta de Castilla-La Mancha y el Fondo Europeo de Desarrollo Regional (FEDER).

Tabla de Contenidos

Lista de Figuras

Lista de Tablas

Acrónimos

AC	**Affective Computing**
ASD	**Autism Spectrum Disorder**
CA	**Computación Afectiva**
HCI	**Human-Computer Interaction**
HERA	**Heterogeneous Emotional Results Aggregator**
IPO	Interacción Persona-Ordenador
TEA	Trastorno del Espectro Autista
TUI	Tangible User Interface
UX	User Experience

Capítulo 1

Introducción

En este capítulo se definirán las bases de esta tesis doctoral. La Sección 1.1 presentará el contexto y circunstancias que motivaron el estudio de las líneas abordadas a lo largo de este trabajo. La Sección 1.2 presenta algunos de los conceptos sobre los que se fundamentan los trabajos acometidos en el contexto de esa tesis. La Sección 1.3 presenta los objetivos principales y los objetivos secundarios que esta tesis pretende satisfacer. Finalmente, se detalla la estructura de este documento en la Sección 1.4.

1.1 Justificación y motivación

El término protagonista de esta tesis, Computación Afectiva (*CA*), fue acuñado por Rosalind W. Picard en 1995 como toda forma de computación "relacionada con, provocada por, o que influye en emociones" [35]. Así, podríamos distinguir diversos tipos de tareas enmarcadas en la *CA* en función de qué se haga con las emociones.

- **Detección**. La detección de emociones es una actividad amplia y multidisciplinar que puede implicar desde el uso de sensores para la medición de señales fisiológicas hasta el entrenamiento de modelos de aprendizaje automático. Esto está directamente relacionado con la manera que tienen las emociones de manifestarse, puesto que, al margen de su definición y/u origen (que puede hacerse desde varias áreas de conocimiento, desde la Psicología a la Filosofía y la Medicina), las emociones tienen una manifestación física, y estas manifestaciones pueden transformarse en datos que un computador puede procesar, traducir y utilizar.

- **Procesamiento**. Según la pureza de nuestra fuente de datos afectivos, el producto de esa detección conllevara un nivel de procesamiento u otro. Se entiende la pureza de un dato como lo cerca que está de convertirse en información afectiva valiosa. Así, un indicador biométrico que indique un nivel de excitación o estrés se considerará más puro que una lista de números que representen el valor RGB de los píxeles de una imagen que contiene un rostro. La explicación de esto es que, para obtener *valor* de ese dato biométrico, apenas habrá que compararlo con un umbral preestablecido para saber qué representa, mientras que para extraer valor de esa lista de números definiendo el color de cada píxel de una imagen habrá que entrenar algún tipo de clasificador o reconocedor de patrones (por ejemplo, una red neuronal) con imágenes de rostros previamente etiquetadas con la emoción que representan.

- **Utilización**. Una vez que una emoción (su manifestación) ha sido transformada en datos almacenados en memoria, esta puede utilizarse de muchas maneras, aunque estos usos pueden distinguirse como usos *pasivos* o *activos*. En el contexto de esta tesis, se denoniman usos *pasivos* a aquellos que no influyen en el usuario directamente y/o a corto plazo. Así, se dice que una aplicación usa las emociones de forma pasiva cuando se limita a registrar las emociones de los usuarios para una posterior visualización, agregación o análisis por parte de un tercero (que podría ser incluso el propio usuario). Por el contrario, se dice que las emociones se usan de forma *activa* cuando se utilizan para influir en el usuario que las origina directamente y/o a corto plazo. Entonces, un ejemplo de aplicación que usase las emociones de forma activa sería aquella que utilizase las emociones detectadas para modificar su comportamiento en tiempo real, para intentar crear una respuesta acorde en el usuario.

- **Simulación**. Por último, si bien esta actividad podría catalogarse como la menos explotada, existe también la posibilidad de simular las emociones. Para ello, se hace uso de algún tipo de avatar con el cual el usuario pueda interactuar y percibir dichas emociones. En este caso, el objetivo principal es despertar cierta *empatía* en el usuario, que al sentir que está interactuando con un ente emocional, tiende a reaccionar de forma distinta a como lo haría si sintiera que solo interactúa con un autómata. Un ejemplo de este tipo de actividad podrían ser el der los asistentes virtuales que algunas paginas webs utilizan para atender a sus clientes, que ofrecen respuestas (desde un cambio en la expresión facial de un avatar 3D a el uso de un vocabulario concreto por parte de un bot de chat) ajustadas al *tono* o *humor* que refleja la entrada del usuario.

Originalmente, el esfuerzo investigador en el campo de la *CA* se volcó en las dos primeras actividades, proliferando trabajos como [21][37][40][42] sobre la detección de emociones usando medios como la cara o la voz, el procesamiento de señales biológicas y su significado emocional, la posible fusión de información afectiva para mejorar la calidad de la información, etc.

Sin embargo, de un tiempo a esta parte ha crecido el interés por explorar las posibilidades de aplicación de la Computación Afectiva. «¿Y si un juego modificase su dificultad dinámicamente para mantener el interés del usuario?». «¿Y si mi aplicación ofreciese ayuda de forma automática cuando se detectase estrés?». «¿Y si mi plataforma registrase el humor de los usuarios mientras consumen contenido y lo utilizasen para recomendar contenido nuevo?». «¿Puede el estado afectivo ayudar a reducir riesgos al conducir?». Estas preguntas, entre otras, son algunas de las cuestiones que muchos investigadores empezaron a plantearse cuando la disciplina conocida como *CA* se asentaba.

Si bien el estudio del estado del arte ha significado el análisis de muchas de las diversas aplicaciones de la *CA*, las áreas que más atención han recibido han sido el aprendizaje virtual (*e-learning*) y la sanidad electrónica (*e-health*), debido a la experiencia previa del grupo de investigación al que pertenece el doctorando.

Es importante resaltar que esta tesis se ha llevado a cabo en el marco de dos proyectos de I+D+i. Por un lado, esta tesis se desarrolló dentro del marco constituido por un proyecto nacional concedido por el Ministerio de Ciencia, Innovación y Universidades con referencia RTI2018–099942-B-I00 y por un proyecto regional con referencia SBPLY/17/180501/000495 concedido por la Junta de Castilla-La Mancha y el Fondo Europeo de Desarrollo Regional (FEDER). Al mismo tiempo, un contrato predoctoral concedido por la Universidad de Castilla-La Mancha para la formación de personal investigador en el marco del Plan Propio de I+D+i, cofinanciado por el Fondo Social Europeo, y publicado en el DOCM con número de anuncio 2018/12504 y número de resolución 2019/451 permitió desarrollar la tesis doctoral de forma exclusiva.

1.2 Estado del arte: conceptos fundamentales y trabajos relacionados

Esta sección presenta una visión general del contexto de la investigación de esta tesis. En primer lugar, se expondrán algunos de los conceptos básicos de la *CA*. En segundo lugar, se expondrán algunos de los trabajos que motivaron el desarrollo de las distintas líneas de investigación de esta tesis.

1.2.1 Conceptos fundamentales relacionados con la Computación Afectiva

Como se introdujo anteriormente, la *CA* fue definida por Rosalind Picard en 1995, momento en que define los términos que sentarían las bases de toda una disciplina, como detección de emociones, expresión de emociones, computadores afectivos o estados afectivos, entre otros. Con el paso de los años, esta disciplina empezaría a ramificarse: detección de emociones en diversos canales, fusión de señales, aprendizaje automático, *deep learning*, modelado conceptual de emociones, estudio de la influencia de las emociones en la experiencia de usuario, etc. Algunos de los hitos o corrientes más destacadas en los últimos años son los siguientes:

- *Exploración de canales afectivos*. Originalmente, se puede considerar que los primeros experimentos en materia de detección de emociones se realizaron atendiendo las manifestaciones físicas más directas de las emociones, como la expresión facial, la voz y las señales fisiológicas de las personas estudiadas [42]. Con el paso de tiempo, los investigadores volvieron la vista a teorías filosóficas que estudiaban las emociones y cómo estas son proyectadas al exterior, así como su influencia en la cognición y comportamiento humanos [6]. Esto dio lugar al estudio de la posible detección de emociones a través de otras manifestaciones como la forma de mover el ratón de un ordenador [47] o la forma de presionar el volante de un coche [32].

- *Fusión de resultados afectivos*. De forma casi contemporánea al estudio de los canales afectivos [41], los investigadores del campo se plantearon la posibilidad de combinar resultados procedentes de canales afectivos distintos para mejorar la precisión y el alcance de las detecciones realizadas, siguiendo los principios de la fusión de señales [21]. En un escenario hipotético, si el rostro de una persona está expresando *alegría* mientras que su voz expresa *tristeza*, el usuario podría estar ocultando sus auténticos sentimientos, de forma consciente o inconsciente. Los detectores de emociones que realizan este tipo de agregación se denominan *detectores multimodales*, viniendo el concepto modalidad de los distintos tipos de canales afectivos que pueden utilizarse para extraer información afectiva: el rostro, la voz, las señales fisiológicas, la postura, el comportamiento, etc. A pesar de ser un tipo de detector superior a los detectores monomodales, la multimodalidad acarrea consigo una miríada de problemáticas que no siempre resulta rentable afrontar.

- *Mejora de precisión de detectores*. Como se ha comentado en el punto anterior, el uso de sistemas multimodales permite mejorar la precisión de las detecciones realizadas: resultados contrarios pueden evidenciar emociones ocultas; resultados alineados permiten validar el resultado final como genuino. No obstante, no hemos de perder de vista la naturaleza de los detectores de emociones, puesto que, en última instancia, estamos trabajando con clasificadores cuyo rendimiento final depende de Los parámetros usados en su configuración, del tipo de modelo usado para crear el clasificador, del tamaño del *dataset* usado para entrarlo, etc. Los avances que se producen a diario en el campo de la Inteligencia Artificial son debidamente estudiados en el campo de la *CA*, siendo muchas las distintas propuestas existentes en términos de modelos [1][2][30], *datasets* [26][28], etc.

- *Aplicación de las emociones detectadas.* Desde el momento en que fue plausible detectar las emociones de usuarios en tiempo real, los investigadores del campo de la Interacción Persona-Ordenador (*IPO*) empezaron a explotar esta información para aprovecharla en el flujo de trabajo de aplicaciones y dispositivos. Disponer de esta información en tiempo real permite disponer de una mayor dimensión de información sobre el usuario, adaptar su experiencia a su estado afectivo, modificar el comportamiento del sistema en base a este para generar un mayor *engagement*, disminuir situaciones de frustración, analizar las reacciones emocionales del usuario ante el sistema a lo largo del tiempo, etc. Debido al carácter multidisciplinar de la *IPO*, la Computación Afectiva empezó a aplicarse en muchos otros campos, como veremos en la siguiente sección.

1.2.2 Trabajos relacionados con la aplicación de la Computación Afectiva

Tal y como se adelantó en la sección anterior, desde el momento en que aparece la detección de emociones como tal, la *IPO* ve en ella un gran potencial para aplicarse a otras áreas. Si bien el concepto de autoadaptación [4] (modificar el comportamiento del sistema de forma automática en base a las emociones del usuario) es, en esencia, sencillo, sus aplicaciones y manifestaciones son muy numerosas.

Cabe destacar, por encima de otros campos, el uso de la *CA* en el campo de la Educación, puesto que es uno de los campos donde la regulación de emociones puede ofrecer más beneficios: está demostrado que cuando se consigue mantener a los estudiantes por debajo de un umbral de estrés y por encima de otro de aburrimiento o desinterés, estos son capaces de alcanzar mayores niveles de concentración, lo que produce también un beneficio a largo plazo en cuestión de retención del conocimiento estudiado [22]. Así, el uso de herramientas que permitan leer las emociones expresadas en el rostro, en la voz, a través de señales fisiológicas, o a través del comportamiento del estudiante, generan una información que podemos utilizar para producir una autorregulación del estado afectivo del estudiante. Si bien esta regulación puede utilizarse en cualquier contexto para mejorar la experiencia de usuario (*user experience*, UX), el campo del *e-learning* ha sido uno de los que más ha podido aprovechar este tipo de tecnología, como podemos ver en la Figura 1.1. Estudios como [4] o [39] recopilan algunos de los trabajos realizados en este campo, en el cual se aprovecha la información afectiva generada por los estudiantes para regular su estado de ánimo durante una clase, durante la realización de ejercicios, etc.

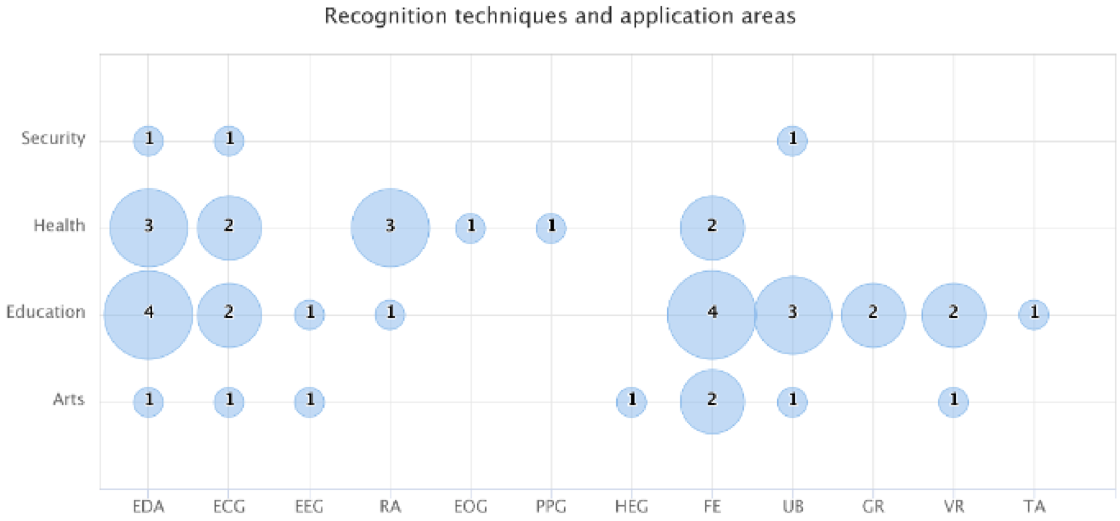


Figura 1.1 Uso de *AC* por área [4]

En segundo lugar, cabe señalar también el campo de la *e-health*, puesto que disponer de las emociones de pacientes durante procesos clínicos puede darnos información valiosa acerca del procedimiento utilizado durante una prueba, acerca de un programa de rehabilitación física de un paciente, favorecer su recuperación o tratamiento, etc. Estudios coetáneos a las primeras propuestas de *CA* existentes ya relacionaban las emociones con la motivación y la velocidad de recuperación de pacientes en procesos de rehabilitación [10], relación que no ha dejado de estudiarse [7][9][38]. En lo que respecta a salud mental, también se han hecho propuestas para integrar la *CA* en procesos de seguimiento y regulación de personas con algún tipo de desorden emocional [23][43], puesto que hacer a los pacientes conscientes de su propio estado de ánimo incrementa su nivel de *awareness*, lo que se traduce en una mejora de sus capacidades de autorregulación.

Es de interés resaltar también un reciente área de estudio que supone la intersección de estas dos últimas áreas mencionadas, y es el estudio y aplicación de la *CA* a personas con Trastorno del Espectro Autista (*TEA*, o *ASD* por sus siglas en inglés). Uno de los obstáculos que afrontan las personas con este tipo de trastorno es la dificultad o imposibilidad de regular sus propias emociones y de identificar las emociones en otros. Sin embargo, con la terapia adecuada (sobre todo en fases iniciales del desarrollo) estas habilidades pueden desarrollarse y entrenarse. Dado que estas intervenciones comienzan en la niñez, es muy común usar la tecnología en forma de *juego serio* para ayudar a los niños a aprender mientras se divierten [48]. Con la llegada de la *CA*, estos juegos empezaron a enriquecerse con habilidades afectivas, permitiendo ampliar el espectro de soluciones ofrecidas a esta problemática [17][31].

1.3 Objetivos

El **objetivo principal** que persigue esta tesis es el de desarrollar nuevos mecanismos de interacción que integren tecnologías propias de la *CA* con técnicas de *IPO* y desarrollar herramientas que den soporte a la creación de dichos mecanismos. Estos mecanismos, a su vez, nos permiten expandir las capacidades de las aplicaciones que las personas usan en su día a día, dotando dichas aplicaciones de la capacidad para leer las emociones de los usuarios, modificar su comportamiento para adaptarse a ellos automáticamente, etc. Este objetivo, a su vez, se ha dividido en una serie de **objetivos específicos**, que son los siguientes.

Cabe señalar que estos objetivos siguen una progresión *bottom-up*, por lo que se parte del estudio de aplicaciones particulares para incrementar el nivel de abstracción en cada paso, culminando con la propuesta de herramientas para el desarrollo y evaluación de los trabajos de los primeros objetivos.

1. **Objetivo 1: revisión del Estado del Arte en el campo de la** *CA*. Estudiar los distintos mecanismos que la *CA* ofrece para crear aplicaciones afectivas, desde la detección automática de emociones hasta su posible inducción o simulación. Esto incluye el estudio multidisciplinar de los distintos canales afectivos conocidos, la exploración de otras fuentes de información afectiva, la comparación de distintas técnicas de clasificación y predicción usando aprendizaje automático, etc.

2. **Objetivo 2: integración de tecnologías CA en sistemas interactivos para mejorar la experiencia de usuario**. Como se presentó anteriormente, la inclusión de información afectiva en el flujo habitual de uso de una aplicación puede mejorar enormemente la experiencia de los usuarios. Es de especial interés su aplicación en el campo del aprendizaje y la salud digitales.

3. **Objetivo 3: desarrollo de prototipos integrando tecnología afectiva para la asistencia en terapias ocupacionales**. Dada la experiencia previa que el grupo de in-

vestigación posee en el campo del aprendizaje y la salud digital, se decidió dar un paso más allá y estudiar posibles coincidencias de estos dos campos y en cómo la *CA* podía aplicarse en ese caso. Un ejemplo de esta triple coincidencia son los procesos terapéuticos con niños con Trastorno del Espectro Autista (TEA), dada su dificultad para reconocer y expresar emociones.

4. **Objetivo 4: definir e implementar una arquitectura de detectores de emociones multimodales que pueda ser adaptada a sistemas distintos con los mínimos cambios**. La detección de emociones tiene un papel central en el desarrollo de aplicaciones afectivas, y la detección multimodal (en más de un canal afectivo a la vez) es reconocida como una forma de detección superior. Sin embargo, esta conlleva unos desafíos que hace que acabe dejándose de lado en favor de detectores de emociones más sencillos, incluso si eso supone usar detectores menos precisos o más propensos a errores de clasificación.

5. **Objetivo 5: propuesta de un modelo de calidad que permita evaluar la calidad de aplicaciones afectivas**. Los modelos de calidad software usados en el mercado actualmente presentan un alto nivel de abstracción, lo que tiende a provocar que no capturen correctamente aspectos más concretos como, por ejemplo, las capacidades afectivas de un sistema o software.

En el siguiente capítulo, revisaremos los resultados que se han obtenido paracada uno de estos objetivos.

1.4 Estructura del documento

Este documento, en el que se presentan las distintas aportaciones y resultados obtenidos en la tesis doctoral, sigue la siguiente estructura.

- En el Capítulo 1 se presenta la justificación y motivación sobre las cuales se sustenta el desarrollo de esta tesis (1.1), el estado del arte en que se enmarca la misma (1.2) y finalmente los objetivos que busca satisfacer (1.3).

- En el Capítulo 2 se muestran las distintas aportaciones y resultados que se han obtenido mientras se abordaba el desarrollo de los objetivos mencionados en la sección 1.3, así como un resumen de los artículos que se han publicado como parte de este compendio (Sección 2.6)

- En el Capítulo 3 se aportan los artículos que forman el compendio de publicaciones de esta tesis doctoral.
- Finalmente, los Capítulos 4 y 5 muestran las conclusiones derivadas del trabajo realizado, así como algunas de las líneas en las que se continuará trabajando en el futuro, tanto en inglés como en español.

Capítulo 2

Aportaciones y resultados obtenidos

Tal y como se introdujo en el capítulo anterior, esta tesis tiene como objetivo principal el desarrollo de nuevos mecanismos de interacción que integren tecnologías propias de la Computación Afectiva con técnicas de Interacción Persona-Ordenador y el desarrollo de herramientas que den soporte a la creación de dichos mecanismos, con el fin de expandir las capacidades de aplicaciones de uso diario con capacidades afectivas. Como se introdujo en el capítulo anterior, este objetivo se dividía a su vez en objetivos específicos. En esta sección, presentaremos los resultados vinculados a cada uno de esos objetivos.

1. **Resultados asociados al objetivo 1: estudio y exploración de los distintos canales afectivos y métodos de detección**. Los distintos trabajos desarrollados durante y antes [15] de esta tesis nos han permitido contemplar el panorama completo en el campo de la detección de emociones. Esto nos ha permitido discretizar los métodos más confiables, los más empleados, los más comercializados, los más sencillos de implementar, propuestas alternativas y/o complementarias a estos, etc.

2. **Resultados asociados al objetivo 2: estudio y exploración de las distintas aplicaciones de la** *CA*. En línea con el resultado anterior, el desarrollo de sistemas que implementen técnicas de *CA* ha significado a su vez la revisión de la literatura en lo que respecta a este menester. Esto nos ha permitido apreciar tanto los trabajos existentes sobre cómo aplicar técnicas de *CA* en todo tipo de aplicaciones como la tendencia existente en este campo.

3. **Resultados asociados al objetivo 3: desarrollo de sistemas interactivos integrando CA**. Para poner a prueba las propuestas derivadas de los estudios sobre aplicaciones de la *CA* se ha llevado a cabo la implementación de varios prototipos que implementen, entre otras funcionalidades, detección de emociones, autorregulación del comportamiento, inducción de estados afectivos en el usuario, etc. Debido a la experiencia

previa del grupo, estos prototipos se enmarcaron en el ámbito del aprendizaje virtual y la sanidad electrónica.

4. **Resultados asociados al objetivo 4: propuesta de nuevas arquitecturas para la detección de emociones**. Tras el desarrollo de algunos prototipos y propuestas sobre distintas formas de detección de emociones, se abordó la propuesta de nuevas arquitecturas para el desarrollo de estos detectores. Si bien la creación de detectores *per se* queda más relegada al campo de la Inteligencia Artificial, la propuesta de arquitecturas para la integración de detectores en otras aplicaciones queda dentro del marco de esta tesis. En concreto, se ha explorado la integración de detectores para la creación de detectores multimodales. Para la validación preliminar de esta propuesta se realizaron pruebas con usuarios que les permitieran apreciar las carencias en el terreno de la combinación de detectores de emociones que nuestra propuesta intentaba solucionar, lo que produjo buenos resultados.

5. **Resultados asociados al objetivo 5: propuesta de un modelo de calidad para valorar aplicaciones afectivas**. Los trabajos realizados en el marco de esta tesis plantearon algunas cuestiones en materia de evaluación y comparación de aplicaciones afectivas. ¿Qué criterios debería utilizar para comparar aplicaciones afectivas, o para evaluar la calidad de una sola aplicación? La ausencia de modelos o métricas específicas para este tipo de aplicaciones nos permitió identificar una laguna en la literatura que intentar corregir.

2.1 Resultados del objetivo 1: revisión del Estado del Arte de la *CA*

Para comenzar a desarrollar el primer objetivo de esta tesis, era necesario analizar aquel elemento que pone en marcha cualquier procedimiento que implique *CA*: las *emociones*. Así, la primera tarea que había de completarse para sentar las bases de este trabajo consistía en estudiar y analizar los canales afectivos que el cuerpo humano utiliza para expresar sus emociones. Como podemos observar en la Figura 2.1, una de las formas de clasificar los canales afectivos es en función de su *visibilidad*. Así, distinguimos entre canales de *manifestación externa*, esto es, aquellos que permiten apreciar el impacto de una emoción «*a simple vista*», como la expresión facial, la voz, la postura, el tamaño de la pupila, etc., y canales de *manifestación interna*, que son aquellos para cuya lectura o detección es necesario un aparato o dispositivo que permita leer señales en el cuerpo humano, como la actividad eléctrica del cerebro, la resistencia eléctrica de la piel, el ritmo cardíaco, la presión sanguínea, etc. Finalmente, merece la pena distinguir un tercer canal de información afectiva que, a pesar de ser algo distinto en su forma de manifestarse respecto a los canales anteriores, también nos permite conocer el estado afectivo de los sujetos a estudio, y es el *estado cognitivo* o comportamiento del sujeto considerado. Y es que dada la relación entre emoción y cognición [34], nos es posible trazar una causalidad entre estados afectivos y estados cognitivos, lo que nos permite apreciar el estado emocional de una persona en su nivel de concentración, los errores que comete, la cantidad de información que retiene, etc.

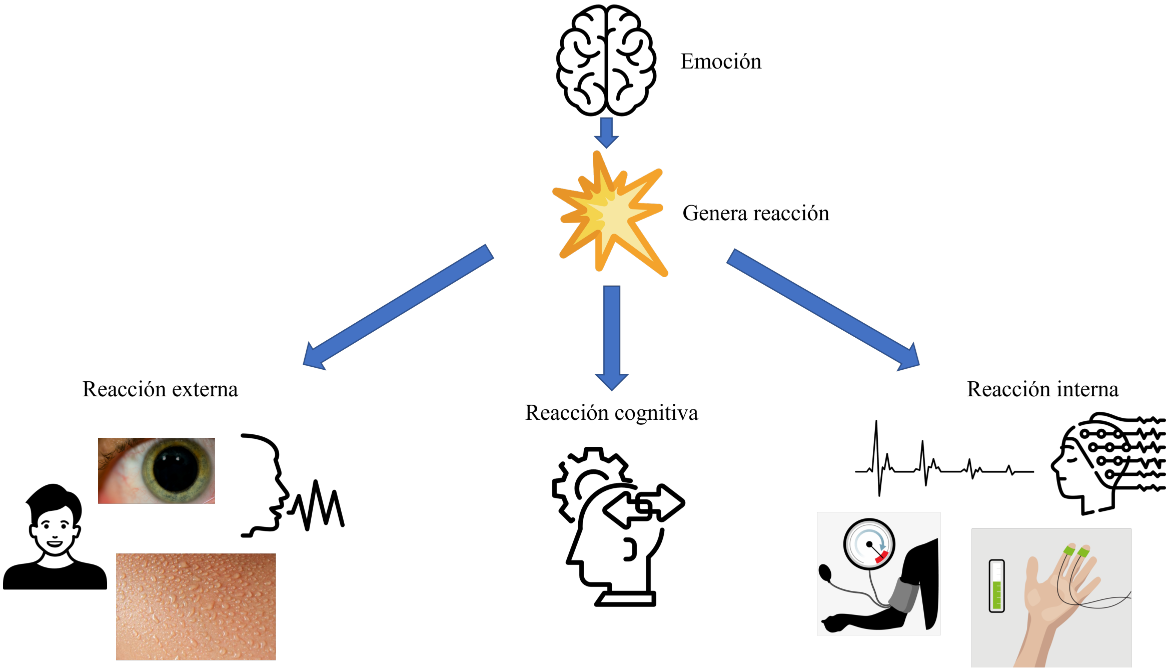


Figura 2.1 Clasificación de canales afectivos

Una vez que hemos analizado los canales afectivos a través de los cuales pueden leerse las manifestaciones de las emociones, hemos de analizar cómo ocurre la detección de emociones. Si bien en el pasado las emociones han sido analizadas desde una óptica incluso mística, para un dispositivo como un ordenador las emociones son algo más físico y material. Así, desde el punto de vista de una máquina, una emoción es una manifestación física medible que la máquina tiene que *aprender* a etiquetar. Si se provee a la máquina de una colección de casos lo suficientemente grande, esta será capaz de clasificar nuevos casos de forma autónoma. Y así es como se adoptan herramientas propias del campo del Aprendizaje Automático para clasificar emociones.

Redes neuronales, modelos ocultos de Markov, máquinas de vectores de soporte, modelos de regresión lineal, modelos mixtos gaussianos: estos son solo algunos ejemplos de las técnicas de clasificación que pueden utilizarse para crear detectores de emociones [45]. Estos modelos son previamente entrenados usando bases de datos afectivas, esto es, colecciones de datos en los que cada registro esta asociado a una emoción. Estos datos pueden ser imágenes de rostros, pistas de audio, señales extraídas de la actividad eléctrica del cerebro o de los músculos, conjuntos de palabras, etc. Tras un proceso de entrenamiento y ajuste, el modelo de clasificación adoptado se vuelve capaz de *clasificar* la emoción que detecte en nuevas consultas que se hagan. En la Figura 2.2 podemos ver un ejemplo de la arquitectura de una red neuronal usada para clasificar emociones en base a expresiones faciales.

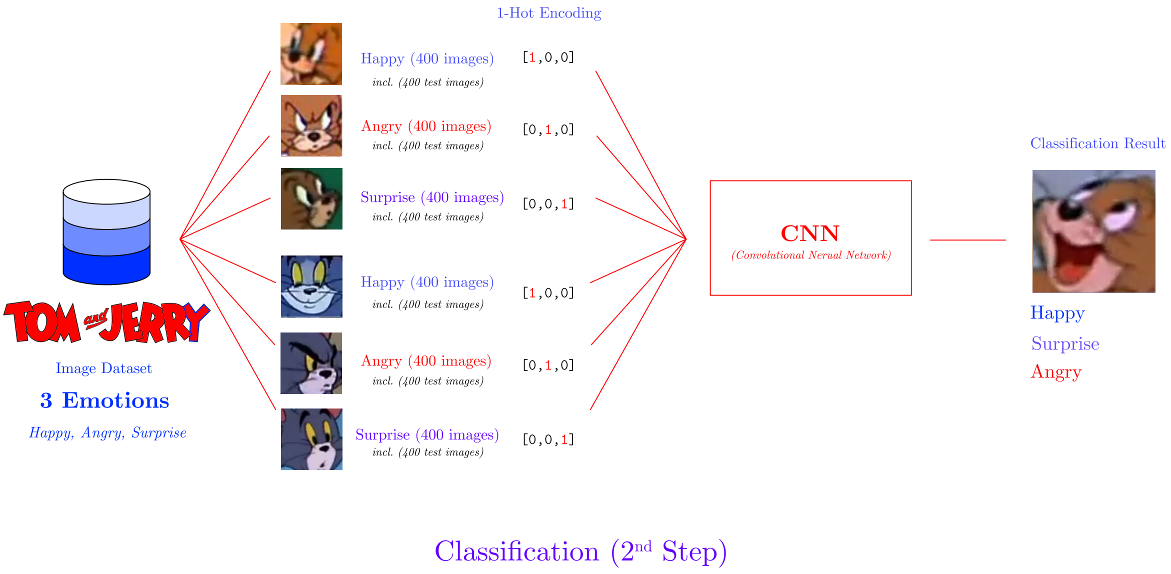


Figura 2.2 Esquema de una red neuronal para clasificar emociones en base a expresiones faciales [24]

Tal y como se puede apreciar en la imagen, la base de datos formada por 2.400 imágenes de rostros de Tom y Jerry se utiliza para entrenar una red neuronal convolucional, red que se vuelve capaz de clasificar nuevas imágenes para indicar si la cara analizada contiene la emoción alegría, sorpresa o enfado.

Como se introdujo en el capítulo anterior, no existe el sistema de detección perfecto, puesto que las características intrínsecas de cada uno de ellos supondrán tanto ventajas como desventajas. Algunas de las características que han de tenerse en cuenta al elegir la detección de emociones que vamos a aplicar son las siguientes:

- *Contexto*. El contexto de uso de un sistema condicionará la detección de emociones que pueda realizarse. Por ejemplo, si el sistema no dispone de capacidad para capturar imagen, no podrá analizarse ni el rostro ni la postura del usuario; si los requisitos exigen que no se utilicen dispositivos invasivos, no podrán leerse valores biométricos, etc.

- *Dificultad de implementación.* Esto incluye tanto la dificultad de desarrollar el detector de emociones al uso como el sistema que utilice dicho detector. La creación del detector supondrá una sobrecarga de tiempo y presupuesto (elegir el clasificador, buscar o crear una base de datos afectiva etiquetada, entrenar el modelo, ajustarlo, etc.), mientras que el desarrollo del sistema en sí supondrá otros desafíos. Algunos ejemplos son la gestión de entradas afectivas, la interconexión con dispositivos biométricos, etc.

- *Responsabilidad del detector.* Tener acceso a un detector propio otorga más control a sus dueños y desarrolladores sobre las detecciones realizadas y elimina la dependencia del servicio ofrecido por un tercero, pero supone una sobrecarga de trabajo adicional (crearlo, entrenarlo, hacerlo accesible, proteger datos de usuarios, etc.). Si en lugar de eso se decide pagar ese servicio de detección a un tercero, el desarrollo se ve agilizado pero añade restricciones asociadas a ese servicio concreto (detecciones disponibles por minuto, dependencia de conexión a Internet, etc.)

- *Presupuesto.* Si bien existen gran cantidad de detectores gratuitos disponibles bajo licencia de código abierto, existe también una gran selección de empresas ofertando servicios de detección de emociones, previo pago de una cuota o de forma gratuita pero restringida. Crear el detector que se necesite puede traducirse en un ahorro en costes a largo plazo, pero eso supone realizar una inversión inicial de recursos para crear el detector. Por el contrario, pagar por el uso del detector de un tercero puede ofrecernos acceso a detectores más perfeccionados, un servicio de atención al cliente, etc.

- *Necesidades de información.* Una vez más, las circunstancias o requisitos del proyecto pueden exigir usar un tipo de detección de emociones sobre otro. Si se necesitan resultados muy fieles, precisos o rápidos, puede que haya que priorizar un tipo de detector o un modelo de detector concreto sobre otro, incluso si esto tiene un impacto en el presupuesto del proyecto.

- *Validez del detector.* Según la fiabilidad del detector elegido y los requisitos de precisión del proyecto en cuestión, puede que sea necesario usar más de un detector del mismo tipo para realizar una doble validación de los resultados, crear un detector multimodal que produzca resultados más fiables o elegir un detector superior.

Las publicaciones y trabajos derivadas de este objetivo son los siguientes:

- Publicación en congreso internacional Interacción 2017 y en ACM Digital Library sobre los distintos canales afectivos a través de los cuales se puede extraer información afectiva, junto con revisión de tecnologías disponibles para realizar dicha extracción [15].

2.2 Resultados del objetivo 2: estudio de aplicaciones de la *CA*

Se puede considerar que los trabajos desarrollados en el contexto de la *CA* han seguido siempre dos direcciones distintas: en primer lugar, el estudio de técnicas de clasificación y aprendizaje automáticos para la detección de emociones, como reflejan las primeras propuestas en materia de *CA* que aparecieron tras su creación [37][40][42]; en segundo lugar, la aplicación de estos nuevos modelos de clasificación y predicción en diversos. La posibilidad de enriquecer el uso de cualquier tipo de sistema nos añade una dimensión de información que podemos usar para moldear la experiencia de usuario y transformarla completamente.

Originalmente, las primeras aplicaciones realizadas apenas eran experimentos ad-hoc sobre aplicar un tipo de tecnología afectiva a un caso de uso particular, especialmente en el campo de la educación, de lo que son prueba trabajos como [5] y [27], desarrollados en el seno de la universidad en la que nació esta disciplina. Más tarde, los desarrolladores de videojuegos con biofeedback vieron en estos una oportunidad para crear videojuegos afectivos. Se denominan *juegos con biofeedback* a aquellos videojuegos en los que la información fisiológica del usuario se utiliza para controlar el juego o influir en él de alguna manera. Así, si esta información fisiológica se utiliza además para influir en el flujo de uso de la aplicación y en el estado afectivo del jugador, ese juego se transforma en un juego afectivo.

Con el paso del tiempo, la *CA* empezó a expandirse a otras áreas de conocimiento, como el marketing [8], la psicología [3], la educación [46], la sanidad [29], etc. Durante una revisión de la literatura realizada en 2019 en la revista *IEEE Transactions on Affective Computing* que incluyó el análisis de 20 artículos, se obtuvo una distribución de disciplinas que podemos observar en la Figura 2.4. En líneas generales, se pudo apreciar un foco más intenso en el estudio de la depresión desde una perspectiva tecnológica así como el análisis del estado cognitivo de las personas, tanto en materia de usabilidad más general como en casos aplicados. En este filtrado de la literatura se apreciaron también algunos trabajos relacionados con el autismo y en como la *CA* podía participar en procesos de terapia conductual.

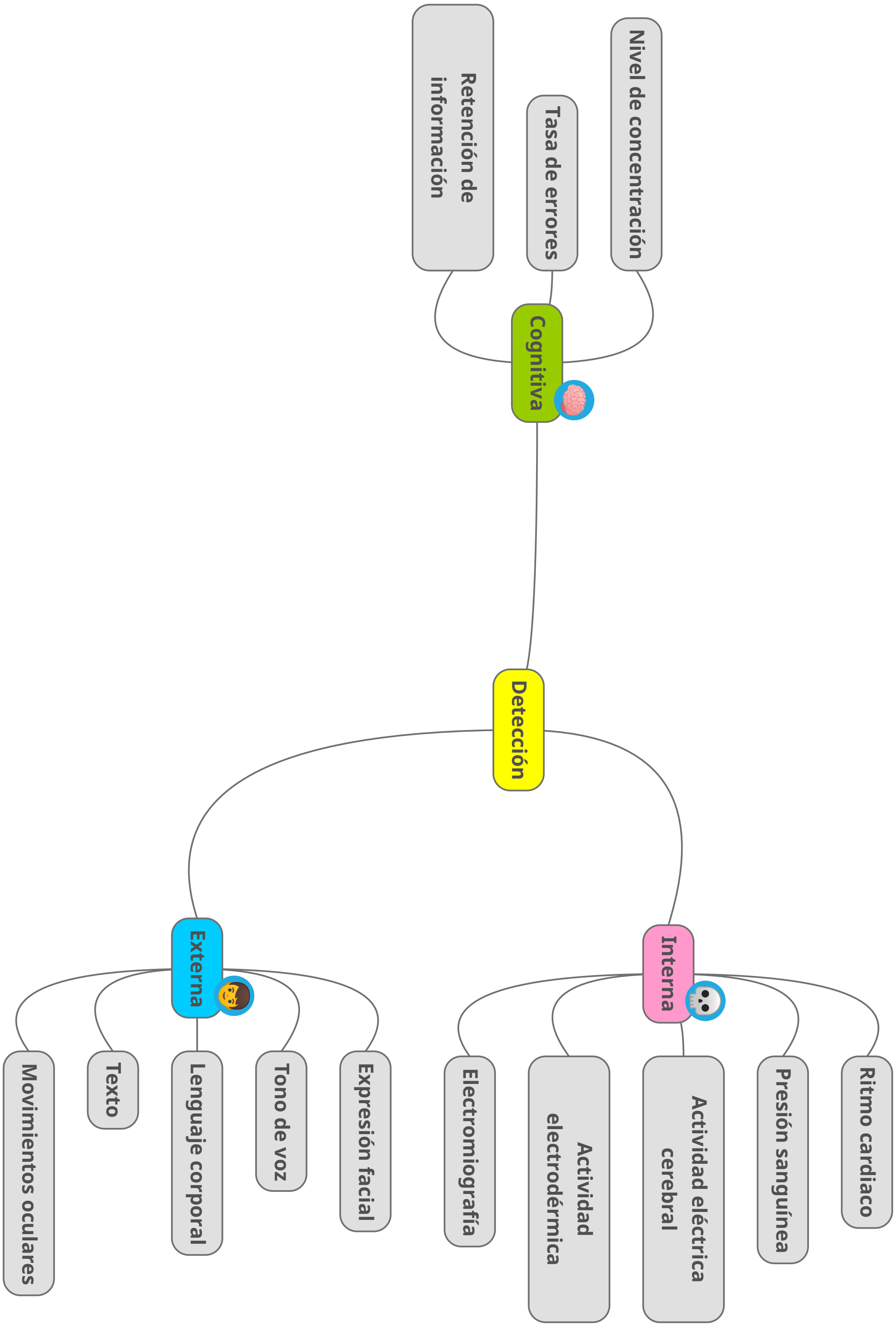


Figura 2.3 Categorización de las formas de detectar emociones

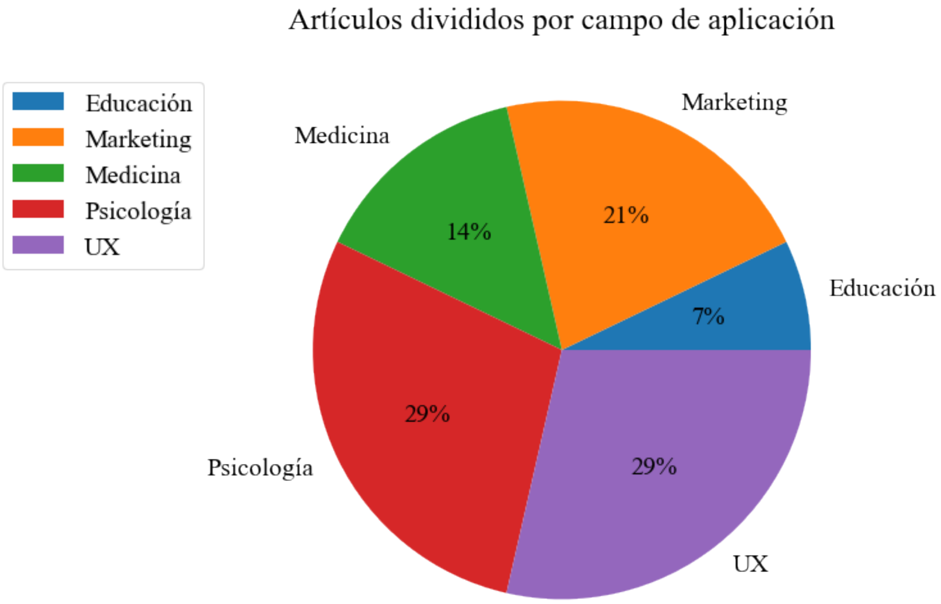


Figura 2.4 Estadística de aplicaciones de la *CA*

En una revisión más reciente [4], se puede apreciar un mayor interés en la aplicación de la *CA* en el terreno educativo, como expuso la Figura 1.1. Esta revisión expuso también una fuerte presencial de la detección de emociones basada en el rostro y en la resistencia eléctrica de la piel.

Las publicaciones y trabajos derivados de este objetivo son los siguientes:

- Publicación en congreso internacional Interacción 2017 sobre los distintos canales afectivos a través de los cuales se puede extraer información afectiva, junto con revisión de tecnologías disponibles para realizar dicha extracción [15].

2.3 Resultados del objetivo 3: desarrollo de prototipos afectivos

Como fruto de las investigaciones iniciadas en el campo de la *CA*, se abordó el desarrollo de diversos prototipos, con el objetivo de aplicar las tecnologías analizadas en los campos que se conocían.

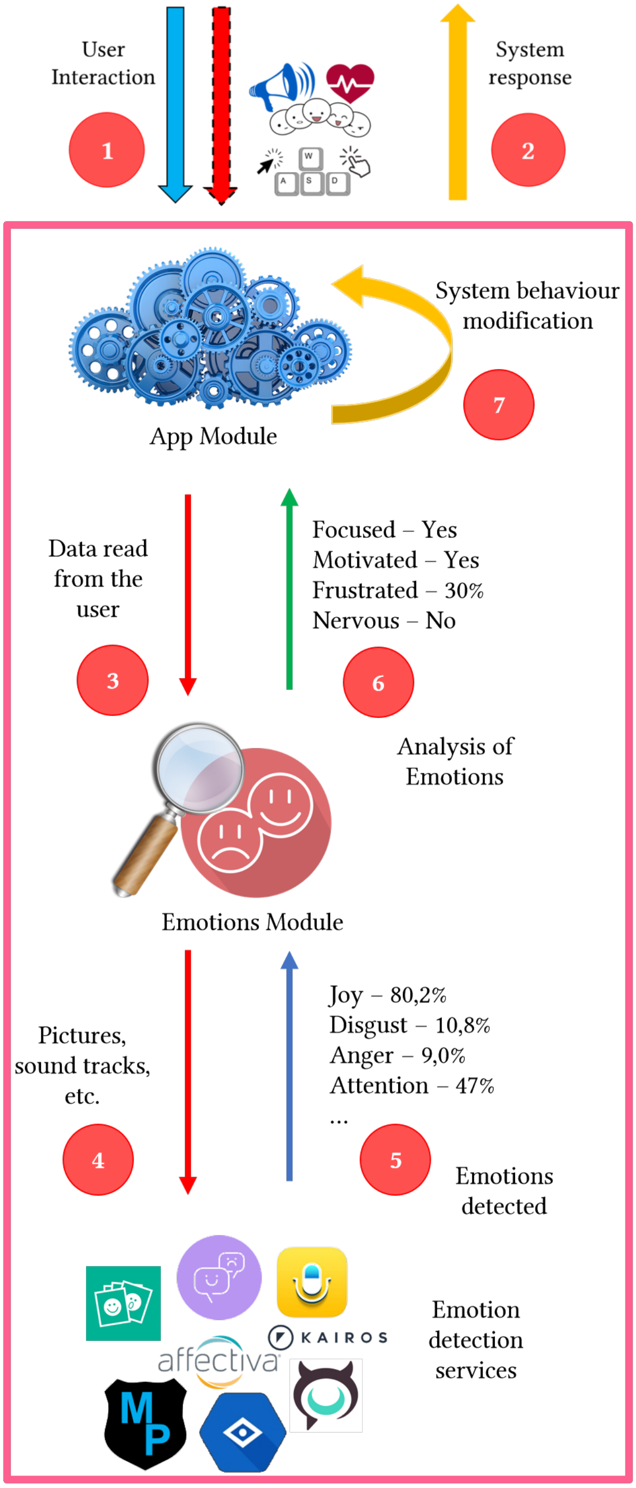


Figura 2.5 Esquema emoCook

El primer trabajo abordado se expone en el siguiente capítulo como uno de los primeros artículos que avalan este compendio y consiste en una aplicación utilizada para impartir inglés a niños españoles de entre 6 y 12 años. Esta aplicación, diseñada como un juego serio, es también una aplicación afectiva con una capacidad de autorregulación, esto es, que utiliza la información afectiva leída para modificar su comportamiento. Dado que el núcleo de la aplicación es un juego, dicha información afectiva se utiliza para modificar la dificultad de este, lo que nos permite desafiar a usuarios más aventajados y animar a progresar a aquellos menos diestros. En la Figura 2.5 podemos ver un esquema de como funciona esta aplicación, emoCook.

emoCook es una aplicación que originalmente fue diseñada para ser usada en un ordenador con teclado, ratón, cámara web y micrófono. Como se puede ver en la figura, el usuario interactúa con el sistema con normalidad (1) y este responde a las entradas del usuario (2). Sin embargo, al mismo tiempo que ocurre esto, el sistema está registrando datos afectivos del usuario (su expresión facial, el tono de su voz y su forma de interactuar con el teclado) y utilizándolos para modificar el comportamiento del sistema. Para realizar este análisis afectivo se utilizaron dos servicios de detección emocional de terceros, *Affectiva* y *Beyond Verbal*. La información derivada de la interacción del teclado se analizaba localmente con un *script* diseñado expresamente para ello. La dificultad se codificó como un valor calculado que controlaba tres parámetros del juego, de manera que los cambios de dificultad fuesen menos *obvios* de cara al usuario.

En segundo lugar se desarrolló EmoTEA. EmoTEA fue diseñado como una aplicación móvil en colaboración con la asociación Autismo Albacete para aunar la terapia de niños con TEA con las nuevas tecnologías afectivas. Es un hecho comprobado que una de las dificultades que enfrentan las personas en el espectro autista es el reconocimiento de emociones de terceras personas, así como la autorregulación de las propias. Sin embargo, ese reconocimiento (tanto externo como interno) puede desarrollarse como una cualidad humana más con las terapias adecuadas [11][12]. En colaboración con la asociación anteriormente mencionada, se abordó el desarrollo de una aplicación para apoyar en los terapeutas y a sus pacientes en el proceso de aprendizaje.

Las dos tecnologías centrales empleadas en EmoTEA son el reconocimiento de emociones basado en la expresión facial y el uso de objetos tangibles (TUI por sus siglas en inglés). El uso de tecnologías de detección de emociones permite el desarrollo de actividades para fomentar el reconocimiento de emociones a través de la *imitación*, puesto que son los propios usuarios, asistidos siempre por el terapeuta, los que aprenderán a expresar sus emociones de forma autónoma. Por otro lado, el uso de objetos tangibles ofrece un mecanismo de interacción mucho más apto para este tipo de terapia que cualquier recurso *point-and-click*

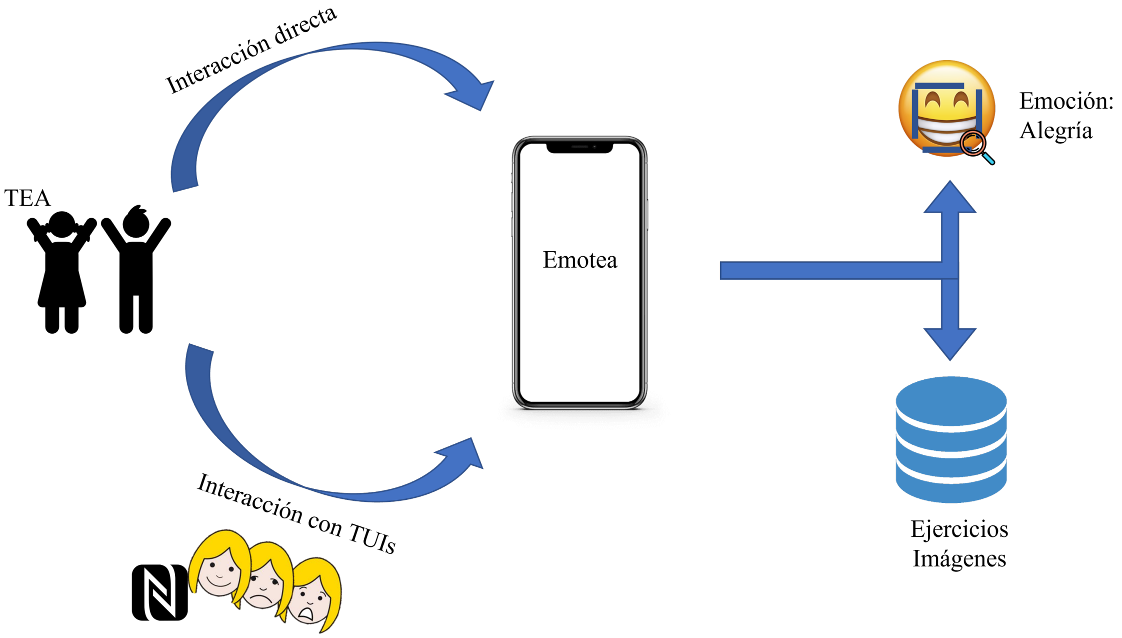


Figura 2.6 Esquema EmoTEA

tradicional. A través de estos objetos tangibles (cartulinas con pictogramas representando emociones y una etiqueta NFC pegada), los niños pueden indicar, en actividades de identificación, que emoción están visualizando. En la Figura 2.6 podemos ver un esquema del flujo de interacción de este prototipo.

A continuación, se dió un paso más allá y se planteó desarrollar una herramienta que pudiera haber sido de ayuda durante las implementaciones previas. Así, se inició la implementación de *HERA* (Heterogeneous Emotional Results Aggregator), una arquitectura presentada en la siguiente sección para el desarrollo de detectores de emoción multimodales. Esta arquitectura ofrece una serie de mecanismos para implementar detectores multimodales, reconocidos como superiores pero también como más complejos, ofreciendo soluciones a los obstáculos que este tipo de detectores suponen. En la Figura 2.7 podemos ver un diagrama que expresa la idea fundamental de este prototipo. Como se puede observar en el esquema izquierdo de la imagen, cuando se quiere combinar distintos tipos de detectores, dadas las propiedades intrínsecas de cada tipo de detección, así como las de los clasificadores usados en cada caso, no es posible combinar los resultados inmediatamente, puesto que sus diferencias, en cuestión de formato, por ejemplo, los hacen incomparables. Para superar este obstáculo, es necesario un paso adicional en el que se haga una comparación manual y totalmente adaptada a los detectores concretos que se estén usando. El objetivo de *HERA* es ofrecer una arquitectura como la que se encuentra en el lado derecho de la imagen, definiendo pasos intermedios de traducción y agregación que permitan obtener un resultado agregado automáticamente.

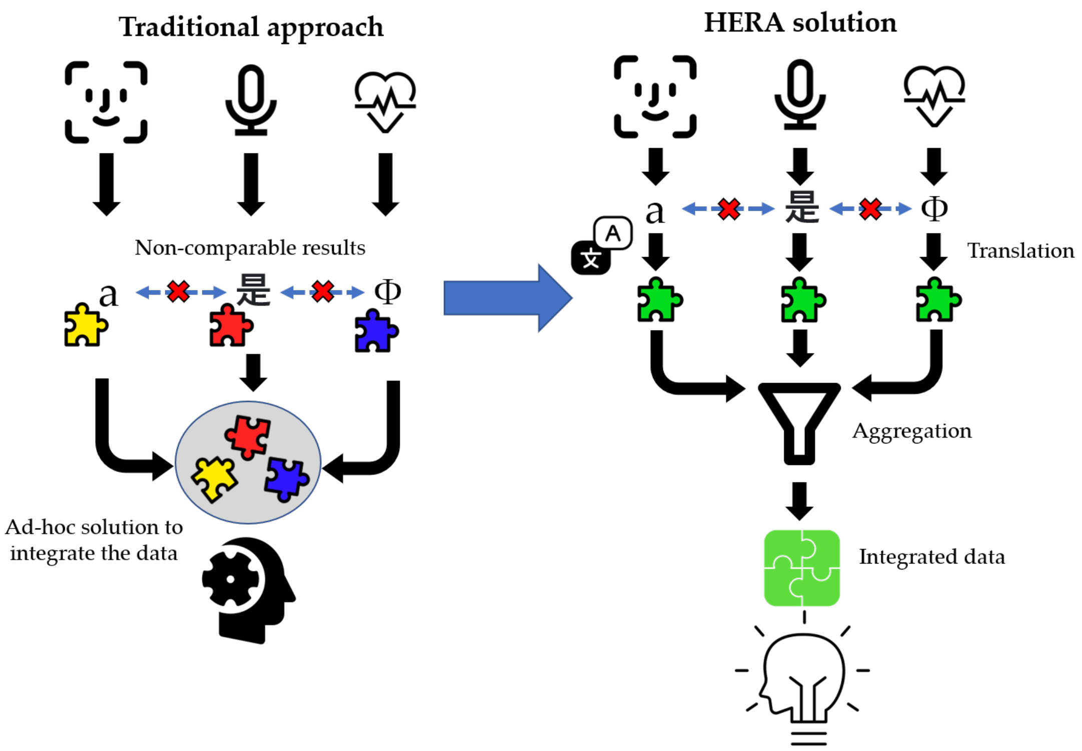
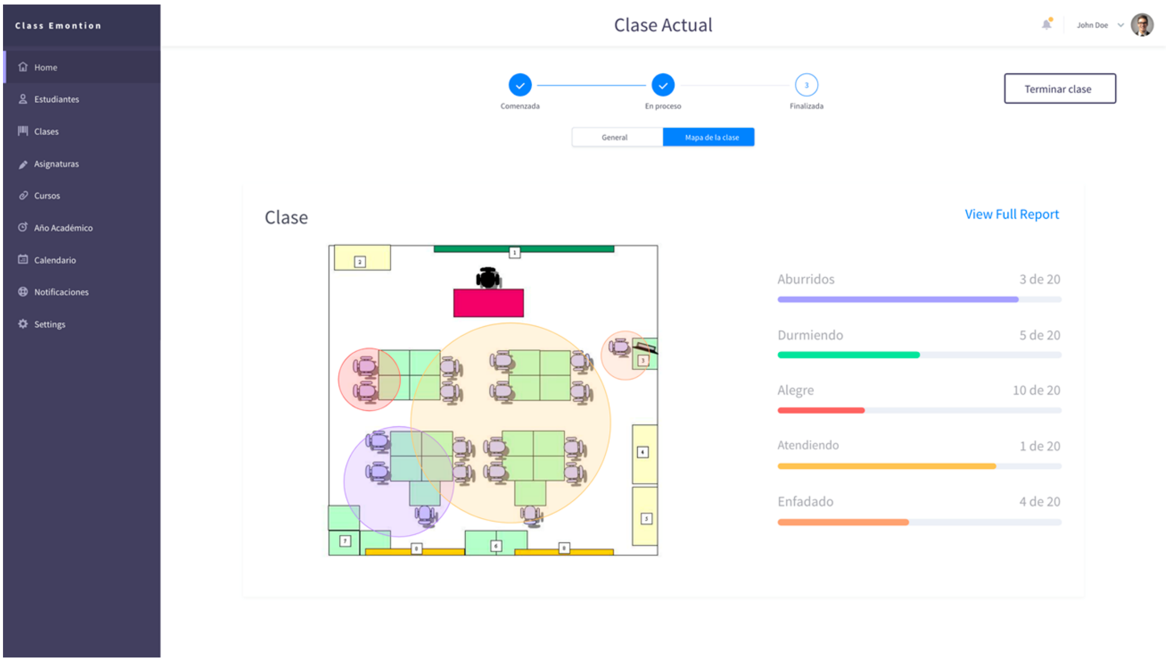


Figura 2.7 Framework HERA

Siguiendo la línea del *e-learning*, se planteó la siguiente cuestión: ¿y si un docente fuera capaz de visualizar, de un solo golpe de vista, el estado afectivo de su clase? Esto le permitiría reconocer ciertos patrones de comportamiento entre estudiantes, identificar estudiantes distraídos o bajo mucho estrés y, en definitiva, disponer de una dimensión más de información sobre sus alumnos. Esta idea se materializó en la forma de EmotionFace, una aplicación desarrollada usando Angular y Node.js que utiliza la cámara web y la pulsera Xiaomi Band 2 para extraer información afectiva de los estudiantes y que permite a sus usuarios docentes monitorizar el estado afectivo de sus alumnos, modificar la distribución de su clase, analizar gráficas y datos históricos, etc. Para ello, los estudiantes se conectan desde sus dispositivos a una sección concreta de la aplicación, donde aceptan que la aplicación acceda a la cámara y a la conexión Bluetooth. Tras pasar el flujo de vídeo de la cámara por un detector de emociones basado en el rostro y registrar los latidos por minuto de cada usuario, la aplicación dispone de información suficiente para mostrar, en una representación esquemática de la clase, el estado afectivo de cada usuario. En la Figura 2.8 podemos ver una idea del mapa de calor que puede mostrar la aplicación al docente mientras se encuentra registrando datos.



Aburridos 3 de 20

Durmiendo 5 de 20

Alegre 10 de 20

Atendiendo 1 de 20

Enfadado 4 de 20

Figura 2.8 Pantalla de control de una clase

En lo que respecta al campo de la salud digital, se comenzó a estudiar el posible papel de las emociones en un proceso de rehabilitación física. Dados los estudios que avalan la relación positiva que existe entre emoción, motivación y velocidad de recuperación [10], se decidió abordar la extensión de la aplicación SIVIRE, una aplicación para realizar ejercicios de telerrehabilitación desde casa usando el dispositivo Kinect v2, para dotarla de capacidades afectivas. Durante la ampliación de este prototipo se mantuvo la funcionalidad central del mismo, esto es, el seguimiento del esqueleto del paciente durante la realización de ejercicios previamente diseñados por fisioterapeutas. Sin embargo, se añadió un servicio de detección de emociones basado en el rostro al prototipo, para que durante la realización de los ejercicios se registrase también el estado afectivo del paciente, muestras de dolor, etc. En la Figura 2.9 podemos ver una pantalla de SIVIRE en la que un usuario está realizando un ejercicio. En esta figura podemos observar, en el lado derecho, un icono que refleja, usando distintos *emoticonos* y colores, la emoción expresada por el paciente durante el último periodo de tiempo. Esta información queda disponible para el fisioterapeuta una vez que ha terminado el ejercicio. La finalidad de esta información es que el fisioterapeuta que lleve el proceso de rehabilitación del paciente en cuestión pueda observar sus niveles emocionales a lo largo de las distintas sesiones, lo que le permita considerar si es necesario modificar los ejercicios, ayudarle en alguno de ellos, etc., procurando que el paciente no se frustre durante el proceso y mantenga su nivel de motivación.

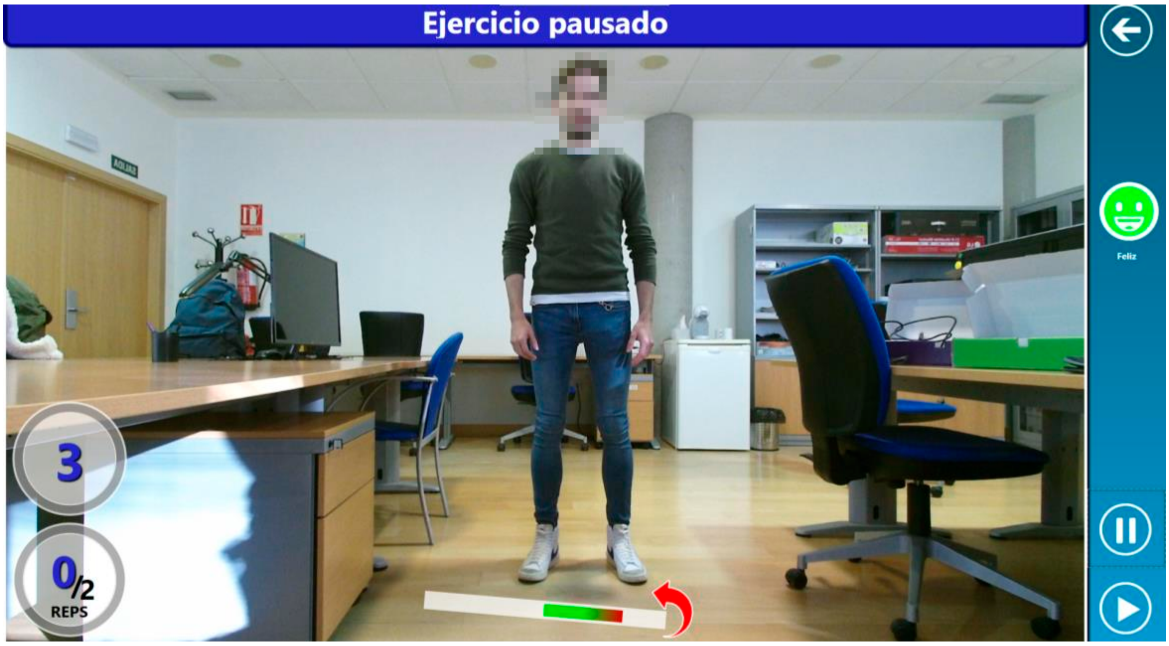


Figura 2.9 Pantalla de realización de ejercicio

Las publicaciones y trabajos derivados de este objetivo son los siguientes:

- Publicación en revista internacional de investigación *Mobile Information Systems* sobre la aplicación de la detección automática de emociones a la autorregulación de juegos serios [16].
- Publicación en congreso internacional *Interacción 2019* sobre la posibilidad de aunar terapias con niños con TEA y detección automática de emociones [17].
- Publicación en revista internacional de investigación *Universal Access in the Information Society* sobre la ampliación de la publicación anterior [19].
- Publicación en revista internacional de investigación *Multimedia Tools and Applications* sobre la implementación de HERA [20].
- Trabajo de Fin de Grado de título "*Detección y visualización de emociones en alumnos en un entorno educativo*", sobre el uso de la detección de emociones en el aula para registrar el nivel de atención de los estudiantes [44].
- Trabajo de Fin de Grado de título "*Mejora por medio de detección de emociones y visualización de estadísticas de una herramienta de rehabilitación basada en movimiento*",

para la mejora de una aplicación de telerrehabilitación, dotándola de capacidades afectivas e implementando un módulo de visualización de datos [13].

2.4 Resultados del objetivo 4: propuesta de arquitecturas para la implementación de detectores

Tras desarrollar los prototipos anteriores, se decidió elevar el nivel de abstracción de la siguiente propuesta, usando la experiencia adquirida en el desarrollo de estos prototipos para identificar necesidades y obstáculos que otros investigadores y desarrolladores trabajando en el campo de la *CA* y la *IPO* pudieran enfrentar. Una de estas necesidades aparece cuando se afronta el desarrollo de detectores multimodales. Si bien su superioridad está reconocida, su desarrollo siempre conlleva una serie de desafíos que incrementan la carga de trabajo de la persona encargada de su creación [33]. La sincronización de resultados, la traducción de estos a un lenguaje común, la correcta agregación de todos ellos teniendo en cuenta distintos factores: todas estas problemáticas han de ser resueltas para poder aprovechar el valor y el poder de los detectores multimodales.

- *Sincronización de servicios*. Todo desarrollo de una aplicación afectiva conllevará la selección de los mecanismos de detección de emociones que vayan a usar. Cuando esta selección se vuelve *múltiple*, recae sobre el desarrollador organizar la detección de emociones de cada canal para que tenga lugar en el momento adecuado, se disponga de todos los resultados individuales a la vez, etc.

- *Agregación de datos heterogéneos*. Si bien la gran mayoría de detectores disponibles en el mercado [15] utilizan un formato similar para expresar sus resultados, no existe ningún estándar que determine cómo expresar la emoción detectada en un recurso analizado. Un detector puede expresar sus resultados usando *categorías* y grados de confianza en que una emoción está presente (`alegría: 80%, neutralidad: 20%`); puede usarse también un conjunto de *dimensiones* que indiquen aspectos de la emoción detectada, como su positividad e intensidad (`positividad: 0.456, intensidad: -0.631`). A su vez esta información puede devolverse en formato JSON, XML, texto plano, etc. Todos estos datos han de integrarse para poder analizarse en el contexto global.

- *Análisis compuesto de resultados*. Una vez recogidos los resultados de los distintos detectores y habiendo traducido estos a un mismo formato para que sean comparables, hemos de *fusionarlos* de forma correcta. Si, por ejemplo, un detector indica que se

> ha detectado alegría con un 100% de confianza, y otro indica que se ha detectado con un 0% de confianza, hacer una media de esos dos valores reduciría ese grado de confianza al 50%, lo que podría conducirnos a una conclusión errónea. Salvo casos excepcionales (imprecisiones, errores del detector), este tipo de conflictos suelen indicar alguna situación más compleja como una ocultación del estado afectivo real por parte del usuario, una fusión incorrecta de resultados afectivos, etc. Si no se hace una fusión que tenga en cuenta los distintos factores que hayan influenciado la producción de esos resultados, podemos acabar perdiendo todos los beneficios de usar detectores multimodales.

Dado el papel central de los detectores en esta arquitectura, el primer paso fue conceptualizar la idea de detector en forma de artefacto. ¿Qué es un detector en el contexto de una aplicación? Se trata de un elemento que es capaz de recibir una entrada (que puede recibir de forma activa o pasiva por parte del usuario) y producir un resultado afectivo. Sin embargo, este elemento puede estar implementado de muchas formas: puede tratarse de un servicio desplegado en un servidor remoto, gestionado por terceras partes, y que recibe solicitudes y envía respuestas usando peticiones HTTP; puede tratarse de un sensor integrado en una pulsera que el usuario lleva puesta y que envía valores biométricos todo el tiempo en forma de paquetes bluetooth a un dispositivo cercano; puede tratarse de una red neuronal que escribe resultados en un fichero. Así, cada detector puede conllevar una inicialización distinta, una forma de petición diferente y un formato específico para generar sus resultados.

Para integrar este tipo de elemento en nuestra arquitectura, diseñamos el artefacto *Detector* (Figura 2.11). Este artefacto implementa una serie de funciones comunes a todos los detectores e incluye tres funcionalidades que han de definirse para cada detector. Así, cada detector ha de implementar un método para llevar a cabo su *inicialización* o conexión, un método para *recibir* la información afectiva y otro método para *traducirla*. Este artefacto es la pieza central de la arquitectura y es usado por el resto de entidades de la misma, como podemos ver en la figura.

De estos pasos mencionados anteriormente, el más importante es el de la traducción, puesto que es el que permite llevar a cabo el resto de pasos de agregación posterior. Como se introdujo anteriormente, no existe un formato aceptado como *estándar* dentro de la comunidad de desarrolladores de aplicaciones afectivas. Sin embargo, si existe un tipo de clasificación de emociones llamada *clasificación dimensional* [6] que permite expresar una emoción en base propiedades como su positividad, intensidad y determinación sentida por el usuario. Este es el caso del esquema de clasificación *PAD* (*Pleasure*, *Arousal* y *Dominance*), en el cual una emoción se representa como un punto en un espacio 3D que indica en qué medida esa emoción es positiva o negativa (*Pleasure*, Valencia), la intensidad de la misma

(*Arousal*, Excitación) y lo dominante o dominado que el usuario se siente al experimentarla (*Dominance*, Dominancia). Este enfoque permite una mayor flexibilidad de cara a la representación de emociones, si bien los detectores usados normalmente tienden a utilizar una *clasificación categórica* en la que la emoción detectada se indica usando las seis emociones universales de Ekman (felicidad, tristeza, sorpresa, miedo, ira y disgusto, con el añadido de neutralidad para indicar la ausencia de emoción) y el grado de confianza o probabilidad con el que cada emoción está presente en la entrada analizada. La adaptación del sistema categórico a un sistema dimensional, como el formato PAD mencionado anteriormente, es el primer paso para poder procesar las emociones provenientes de distintas fuentes (Figura 2.10).

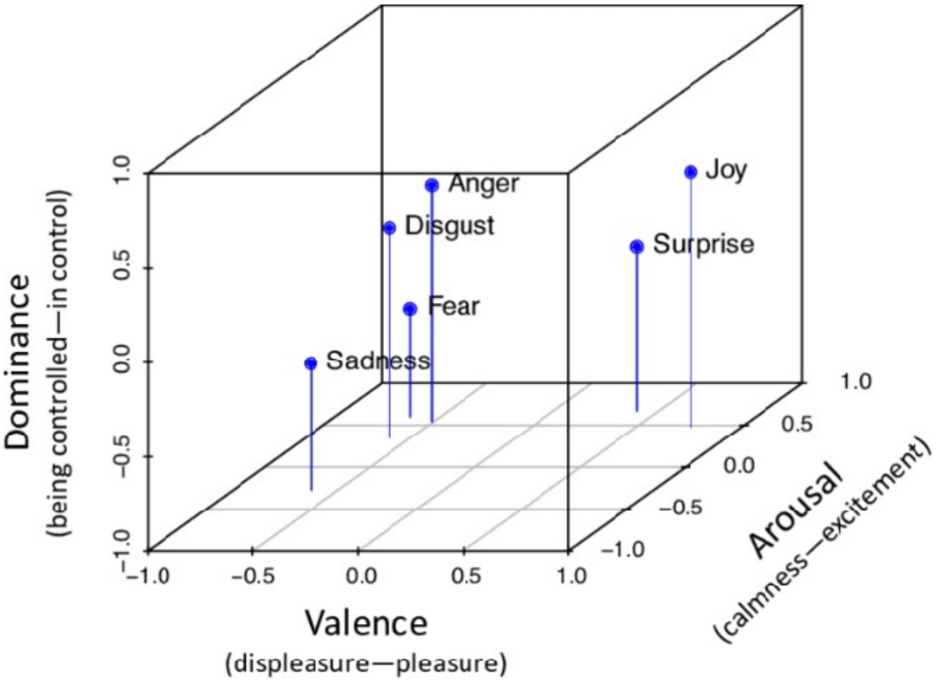


Figura 2.10 Correspondencia entre espacio categorico y espacio PAD ([14])

Dicha traducción ha de ser adaptada para cada detector, debido a las idiosincrasias propias de cada tipo de detección. Así, cada formato particular debe traducirse, mediante las operaciones pertinentes, a una tripla de tres valores. En la gran mayoría de casos esto solo requiere una transformación matemática para trasladar un valor o un conjunto de valores en una tripla, para lo cual pueden usarse transformaciones como las propuestas en [14] u otras derivadas del estudio empirico de cada detector. La fortaleza de esta propuesta reside en su extendibilidad, por lo que se vuelve sencillo, con el tiempo, disponer de una base de transformaciones que den soporte a un gran número de detectores, y de igual forma sucede con las estrategias.

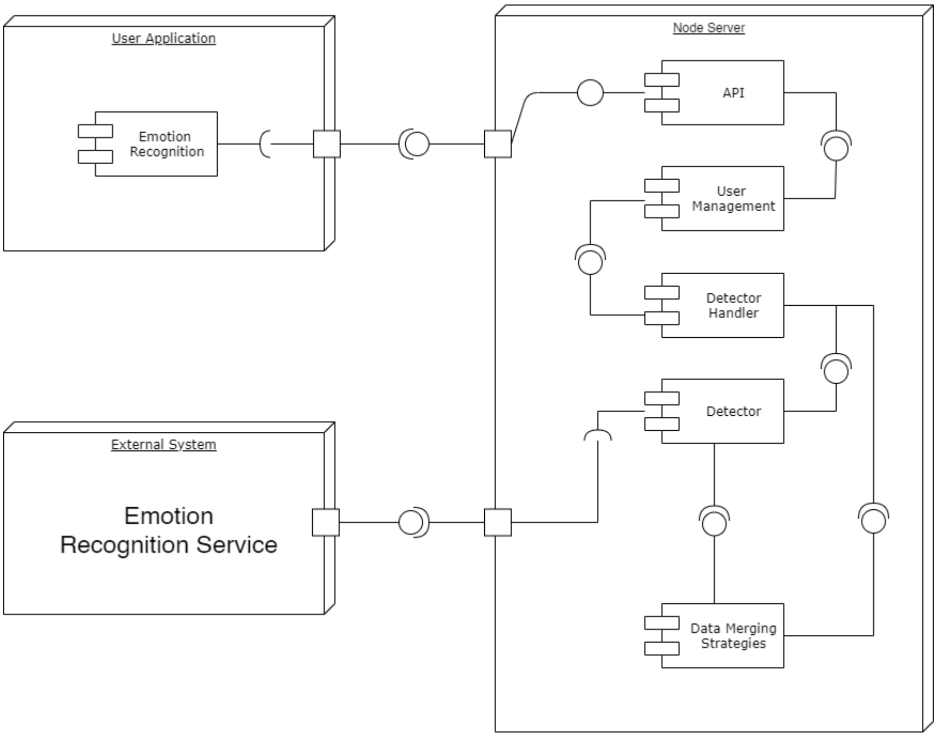


Figura 2.11 Diagrama de despliegue de la arquitectura propuesta

Una vez que se ha realizado la traducción de todos los valores individuales, podemos comenzar su agregación mediante el uso de distintas estrategias. En el contexto de esta arquitectura, una *estrategia* es cualquier operación que recibe una una lista de resultados afectiva y las concentra en un resultado único. Las estrategias pueden ser de cualquier tipo y realizar cualquier operación siempre que respeten ese flujo de entrada y salida. Estas estrategias son usadas una vez por detector, una vez por canal de detección y una última vez para agregar esos resultados en un único resultado final. Este resultado final puede enriquecerse en cada fase para aumentar su valor, siempre respetando que contenga una tripla de tres valores.

Esta arquitectura, implementada usando JavaScript y Node.js con el nombre de *HERA* (*Heterogeneous Emotional Results Aggregator*) implementa una serie de detectores y estrategias para ofrecer una prueba de concepto de esta implementación. Gracias a la flexibilidad de esta arquitectura, HERA puede enriquecerse para dar soporte a mayor tipo de detectores y estrategias mediante la adición de nuevos ficheros implementando cada uno de estos elementos. Para la adición de detectores solo es necesario añadir nuevos ficheros con extensión `js` que implementen las tres funcionalidades básicas mencionadas anteriormente. Esta implementación ha de incluir una función de inicialización, una función de extracción de información emocional y una función de traducción. Una vez más, esta traducción ha de ser implementada según las características de ese nuevo detector. Para el caso de las estrategias, solo hay que respetar el formato de entrada y salida mencionado anteriormente.

Las publicaciones y trabajos derivados de este objetivo son los siguientes:

- Publicación en revista internacional de investigación *Multimedia Tools and Applications* sobre la implementación de *HERA* [20].

2.5 Resultados del objetivo 5: propuesta de un modelo de calidad afectivo

Por último, afrontamos el estudio de la calidad de aplicaciones afectivas. En lo que respecta a la medición de la calidad del software, la ISO 25010 del conjunto de normas SQuaRE ofrece un modelo de calidad para medir la calidad interna, externa y en uso de cualquier tipo de software, proponiendo características, subcaracterísticas y métricas que es interesante medir para conocer la calidad de un software [25]. Sin embargo, muchas de estas métricas obvian problemáticas y características propias de la aplicaciones afectivas. Es por esto que, para superar esta problemática, analizamos distintas estrategias para la extensión y/o propuesta de modelos de calidad. Siguiendo estrategias de extensión aceptadas en la comunidad [36], y partiendo de nuestra propia experiencia con aplicaciones de esta naturaleza, llevamos a cabo un proceso *bottom-up* para la propuesta de nuestra extensión:

1. *Identificar medidas base*. Partiendo de evaluaciones realizadas y/o software desarrollado, identificar métricas que es interesante medir y que influyen en la calidad del sistema afectivo a estudio.

2. *Identificar medidas derivadas*. Analizar las medidas base en conjunto para expresarlas, si es pertinente, como medidas compuestas.

3. *Identificar indicadores*. Aunar medidas derivadas y/o medidas base para generar indicadores, esto es, valores de mayor nivel que dan ideas más generales acerca de la calidad o comportamiento de la aplicación estudiada.

4. *Identificar posibles relaciones entre medidas*. Esto implica detallar cómo se obtienen las métricas y con que otras métricas o características están relacionadas.

5. *Definir subcaracterísticas y características que agrupen las medidas propuestas*. Siguiendo la estructura en árbol usada en la ISO 25010, agrupar características, subcaracterísticas y medidas propuestas en un solo modelo.

Finalmente, y como un recurso de prevalidación, se relacionan las medidas y características propuestas con el modelo de calidad vigente como una forma de remarcar la cualidad de extensión del nuevo modelo, si bien se ha realizado un ejercicio de validación aplicando el modelo propuesto a aplicaciones que no se consideraron cuando se diseñó, lo que también nos permite medir su flexibilidad y completitud: ¿cómo se adapta el modelo a un tipo de software que no se consideró durante su diseño? ¿Permite este estudiar todas las cualidad de este nuevo software o no es capaz de adaptarse a software nuevos? La riqueza de los datos obtenidos produjo conclusiones muy interesantes, si bien se realizarán más ejercicios de validación en el futuro.

En la Tabla 2.1 podemos ver esa correspondencia entre el modelo vigente y nuestra propuesta, que evidencia que el uso de tecnologías afectivas tiene un impacto mayor en el rendimiento, la funcionalidad y la usabilidad.

Tabla 2.1: Equivalencia de modelo de calidad de ISO 25010 y modelo propuesto

SQuaRE Norm		**AffectiveQM**	
Characteristic	**Sub-characteristic**	**Proposed metric**	**Proposed sub-characteristic**
Functional suitability	Appropiateness	Raw data persistence	Device interoperability
		Data persistence	Emotion detection logistics
	Accuracy	Devices accuracy	Device interoperability
		Type of physical manifestation	Interaction mechanism
	Compliance	–	–
Reliability	Availability	Number of extracting stations	Device interoperability
		Devices available	Device interoperability
		Number of affective channels	Emotion detection logistics
	Fault tolerance	Maintenance difficulty	Device interoperability
		Number of extracting stations	Device interoperability
		Reading error ratio	Device interoperability
		Owned services	Usage of Affective Software
	Recoverability	–	–
	Compliance	–	–

Continúa en la siguiente página

Tabla 2.1: Equivalencia de modelo de calidad de ISO 25010 y modelo propuesto (Continuación)

SQuaRE Norm		**AffectiveQM**	
Performance efficiency	Time behaviour	Time to analyse media resources	Emotion detection logistics
		Time to adapt the system	Emotion detection logistics
		Synchronicities overload	Service simultaneity
		Maximum wait time	Service simultaneity
	Resource utilization	Maintenance difficulty	Device interoperability
		Raw data generation	Device interoperability
		Raw data efficiency	Device interoperability
		Number of sensors reading simultaneously	Device interoperability
		Raw data postprocessing difficulty	Device interoperability
		Detectors available	Usage of Affective Software
		Data generation	Emotion detection logistics
		Data efficiency	Emotion detection logistics
		Data persistence	Emotion detection logistics
		Processing location load	Emotion detection logistics
		Number of local services	Usage of Affective Software
		Number of remote services	Usage of Affective Software
		Cost of services	Usage of Affective Software
		Maximum requests per minute	Usage of Affective Software
		Average requests per minute	Usage of Affective Software
		Responses postprocessing difficulty	Emotion detection logistics
		Optimized synchronicities	Service simultaneity
	Compliance	–	–
Security	Confidentiality	Security of requests	Usage of Affective Software
	Integrity	–	–

Continúa en la siguiente página

Tabla 2.1: Equivalencia de modelo de calidad de ISO 25010 y modelo propuesto (Continuación)

SQuaRE Norm		AffectiveQM	
	Non-repudiation	–	–
	Accountability	–	–
	Authenticity	–	–
	Compliance	–	–
Security in use	Operator health and safety	Devices intrusiveness	Device interoperability
	Public health and safety	–	–
	Environmental harm in use	–	–
	Commercial damage in use	–	–
	Compliance	–	–
Usability in use	Effectiveness in use	Type of adaptations	Device interoperability
		Errors in an interaction channel	Interaction mechanisms
		Cost of the services	Usage of Affective Software
	Efficiency in use	User efficiency	Emotion detection logistics
		Devices intrusiveness	Device interoperability
	Satisfaction in use	Type of adaptations	Emotion detection logistics
		User efficiency	Emotion detection logistics
	Compliance	–	–
Flexibility in use	Accessibility in use	Number of interaction channels	Emotion detection logistics
	Context conformity in use	–	–

Continúa en la siguiente página

Tabla 2.1: Equivalencia de modelo de calidad de ISO 25010 y modelo propuesto (Continuación)

SQuaRE Norm		AffectiveQM	
	Context extendibility in use	–	–
	Compliance	–	–

Las publicaciones y trabajos derivadas de este objetivo son los siguientes:

- Publicación en proceso de envío a revista *Computer Standards & Interface* sobre la ampliación del modelo de calidad propuesto en las normas SQuaRE para su adecuación a la evaluación de aplicaciones afectivas.

2.6 Resumen de los artículos publicados y enviados

La Tabla 2.2 muestra la lista de artículos publicados y enviados (en proceso de revisión), como resultado de esta tesis. Nótese que las publicaciones que conforman esta tesis por compendio están resaltadas con un asterisco antes del título. Para cada publicación, se indica también su relación con los distintos objetivos planteados.

Tabla 2.2 Principales publicaciones llevadas a cabo durante la tesis

Publicación	**Objetivos**
Emotion detection: a technology review (Garcia-Garcia, J. M., Penichet, V. M. R., Lozano, M. D.) (2017) (Interacción 2017, ACM Digital Library)	1, 2
* **Multimodal Affective Computing to Enhance the User Experience of Educational Software Applications** (Garcia-Garcia, J. M., Penichet, V. M. R., Lozano, M. D., Garrido, J. E., Lai-Chong Law, E) (2018) (Mobile Information Systems, **JCR-Q3**)	3
Emotea: Teaching children with autism spectrum disorder to identify and express emotions (Garcia-Garcia, J. M., Cabañero, Maria del Mar, Penichet, V. M. R., Lozano, M. D.) (2019) (Interacción 2019, ACM Digital Library)	3
* **Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions** (Garcia-Garcia, J. M., Penichet, V. M. R., Lozano, M. D., Fernando, A.) (2021) (Universal Access in the Information Society, **JCR-Q3**)	3
* **Building a three-level multimodal emotion recognition framework** (Garcia-Garcia, J. M., Lozano, M. D., Penichet, V. M. R., Law, E. L.-C) (2022) (Multimedia Tools and Applications, **JCR-Q2**)	3, 4
Extending the SQuaRE Norms for the Quality Assessment of applications involving Affective Computing (Garcia-Garcia, J. M., Lozano, M. D., Penichet, V. M. R., Kristensson, Per Ola) (2023) (Computer Standards & Interfaces, **JCR-Q1**)	5

Capítulo 3

Publicaciones

En esta sección se presentan las publicaciones que forman esta tesis por compendio, sin perjuicio de otros artículos que se han producido durante la misma. En las siguientes subsecciones podemos encontrar los artículos de revista que conforman este compendio.

3.1 Multimodal affective computing to enhance the user experience of educational software applications

Citación: Garcia-Garcia, J. M., Penichet, V. M. R., Lozano, M. D., Garrido, J. E., and Law, E. L.- C. (2018). *Multimodal affective computing to enhance the user experience of educational software applications*. Mobile Information Systems, ISSN: 1875-905X. Volumen 2018, páginas 10. https://doi.org/10.1155/2018/8751426

Índice de impacto: 1.863, Q3 (JCR 2018)

Esta publicación avala esta tesis por compendio de publicaciones.

Hindawi
Mobile Information Systems
Volume 2018, Article ID 8751426, 10 pages
https://doi.org/10.1155/2018/8751426

Research Article

Multimodal Affective Computing to Enhance the User Experience of Educational Software Applications

Jose Maria Garcia-Garcia,[1] **Víctor M. R. Penichet**,[1] **María Dolores Lozano**,[1] **Juan Enrique Garrido**,[2] **and Effie Lai-Chong Law**[3]

[1] *Research Institute of Informatics, University of Castilla-La Mancha, Albacete, Spain*
[2] *Escuela Politécnica Superior, University of Lleida, Lleida, Spain*
[3] *Department of Informatics, University of Leicester, Leicester, UK*

Correspondence should be addressed to María Dolores Lozano; maria.lozano@uclm.es

Received 19 January 2018; Revised 8 July 2018; Accepted 5 August 2018; Published 13 September 2018

Academic Editor: Salvatore Carta

Affective computing is becoming more and more important as it enables to extend the possibilities of computing technologies by incorporating emotions. In fact, the detection of users' emotions has become one of the most important aspects regarding Affective Computing. In this paper, we present an educational software application that incorporates affective computing by detecting the users' emotional states to adapt its behaviour to the emotions sensed. This way, we aim at increasing users' engagement to keep them motivated for longer periods of time, thus improving their learning progress. To prove this, the application has been assessed with real users. The performance of a set of users using the proposed system has been compared with a control group that used the same system without implementing emotion detection. The outcomes of this evaluation have shown that our proposed system, incorporating affective computing, produced better results than the one used by the control group.

1. Introduction

In 1997, Rosalind W. Picard [1] defined Affective Computing as "computing that relates to, arises from, or influences emotions or other affective phenomena." Since then, a general concern about the consideration of the emotional states of users for different purposes has arisen in different research fields (phycology [2, 3], marketing, computing, etc.).

Concretely, the underlying idea of Affective Computing is that computers that interact with humans need the ability to at least recognize affect [4]. Indeed, affective computing is a new field, with recent results in areas such as learning [5], information retrieval, communications [6], entertainment, design, health, marketing, decision-making, and human interaction where affective computing may be applied [7]. Different studies have proved the influence of emotions in consumers' behaviour [8] and decision-making activities [9].

In computer science research, we could study emotions from different perspectives. Picard mentioned that if we want computers to be genuinely intelligent and to interact naturally with us, we must give computers the ability to recognize, understand, even to have and express emotions. In another different research work, Rosalind pointed out some inspiring challenges [10]: sensing and recognition, modelling, expression, ethics, and utility of considering affect in HCI. Studying such challenges still makes sense since there are gaps to be explored behind them. In human-computer interaction, emotion helps regulate and bias processes in a helpful way.

In this paper, we focus our research in the use of emotions to dynamically modify the behaviour of an educational software application according to the user feelings, as described in Section 3. This way, if the user is tired or stressed, the application will decrease its pace and, in some cases, the level of difficulty. On the other hand, if the user is getting bored, the application will increase the pace and the difficulty level so as to motivate the user to continue using the application.

Finally, we have assessed the application to prove that including emotion detection in the implementation of educational software applications considerably improves users' performance.

The rest of the paper is organized in the following sections: In Section 2, some background concepts and related works are presented. In Section 3, we describe the educational software application we have developed enhanced with affective computing-related technologies. Section 4 shows the evaluation process carried out to prove the benefits of the system developed. Finally, Section 5 presents some conclusions and final remarks.

2. Background Concepts and Related Works

In this section, a summary of the background concepts of affective computing and related technologies is put forward. We provide a comparison among the different ways of detecting emotions together with the technologies developed in this field.

2.1. Affective Computing. Rosalind Picard used the term "affective computing" for the first time in 1995 [11]. This technical report established the first ideas on this field. The aim was not to answer questions such as "what are emotions?," "what causes them?," or "why do we have them?," but to provide a definition of some terms in the field of affective computing.

As stated before, the term "affective computing" was finally set in 1997 as "computing that relates to, arises from, or deliberately influences emotion or other affective phenomena" [1]. More recently, we can find the definition of Affective computing as the study and development of systems and devices that can recognize, interpret, process, and simulate human affects [4]. In other words, any form of computing that has something to do with emotions. Due to the strong relation with emotions, their correct detection is the cornerstone of Affective Computing. Even though each type of technology works in a specific way, all of them share a common core in the way they work, since an emotion detector is, fundamentally, an automatic classifier.

The creation of an automatic classifier involves collecting information, extracting the features which are important for our purpose, and finally training the model so it can recognize and classify certain patterns [12]. Later, we can use the model to classify new data. For example, if we want to build a model to extract emotions of happiness and sadness from facial expressions, we have to feed the model with pictures of people smiling, tagged with "happiness" and pictures of people frowning, tagged with "sadness." After that, when it receives a picture of a person smiling, it identifies the shown emotion as "happiness," while pictures of people frowning will return "sadness" as a result.

Humans express their feelings through several channels: facial expressions, voices, body gestures and movements, and so on. Even our bodies experiment visible physical reactions to emotions (breath and heart rate, pupil's size, etc.).

Because of the high potential of knowing how the user is feeling, this kind of technology (emotion detection) has experienced an outburst in the business sector. Many technology companies have recently emerged, focused exclusively on developing technologies capable of detecting emotions from specific input. In the following sections, we present a brief review of each kind of affective information channel, along with some existing technologies capable of detecting this kind of information.

2.2. Emotion Detection Technologies. This section presents a summary of the different technologies used to detect emotions considering the various channels from which affective information can be obtained: emotion from speech, emotion from text, emotion from facial expressions, emotion from body gestures and movements, and emotion from physiological states [13].

2.2.1. Emotion from Speech. The voice is one of the channels used to gather emotional information from the user of a system. When a person starts talking, they generate information in two different channels: primary and secondary [14].

The primary channel is linked to the syntactic-semantic part of the locution (what the person is literally saying), while the secondary channel is linked to paralinguistic information of the speaker (tone, emotional state, and gestures). For example, someone says "That's so funny" (primary channel) with a serious tone (secondary channel). By looking at the information of the primary channel, the message received is that the speaker thinks that something is funny and by looking at the information received by the secondary channel, the real meaning of the message is worked out: the speaker is lying or being sarcastic.

Four technologies in this category can be highlighted: *Beyond Verbal* [15], *Vokaturi* [16], *EmoVoice* [17] and *Good Vibrations* [18]. Table 1 shows the results of the comparative study performed on the four analyzed technologies.

2.2.2. Emotion from Facial Expressions. As in the case of speech, facial expressions reflect the emotions that a person can be feeling. Eyebrows, lips, nose, mouth, and face muscles: they all reveal the emotions we are feeling. Even when a person tries to fake some emotion, still their own face is telling the truth. The technologies used in this field of emotion detection work in an analogous way to the ones used with speech: detecting a face, identifying the crucial points in the face which reveal the emotion expressed, and processing their positions to decide what emotion is being detected.

Some of the technologies used to detect emotions from facial expressions are *Emotion API* (Microsoft Cognitive Services) [19], *Affectiva* [20], *nViso* [21], and *Kairos* [22]. Table 2 shows a comparative study.

As far as the results are concerned, every tested technology showed considerable accuracy. However, several conditions (reflection on glasses and bad lightning) mask important facial gestures, generating wrong results. For

Table 1: Comparison of emotion detection technologies from speech.

Name	API/SDK	Requires Internet	Information returned	Difficulty of use	Free software
Beyond verbal	API	Yes	Temper, arousal, valence, and mood (up to 432 emotions)	Low	No
Votakuri	SDK	No	Happiness, neutrality, sadness, anger, and fear	Medium	Yes
EmoVoice	SDK	No	Determined by developer	High	Yes
Good vibrations	SDK	—	Happy level, relaxed level, angry level, scared level, and bored level	Medium	No

Table 2: Comparison of emotion detection technologies from facial expressions.

Name	API/SDK	Requires Internet	Information returned	Difficulty of use	Free software
Emotion API	API/SDK	Yes	Happiness, sadness, fear, anger, surprise, neutral, disgust, and contempt	Low	Yes (limited)
Affectiva	API/SDK	Yes	Joy, sadness, disgust, contempt, anger, fear, and surprise[1]	Low	Yes, with some restriction
nViso	API/SDK	No	Happiness, sadness, fear, anger, surprise, disgust, and neutral	—	No
Kairos	API/SDK	Yes	Anger, disgust, fear, joy, sadness, and surprise[2]	Low	Yes, only for personal use

[1]Besides, it also detects different facial expressions, gender, age, ethnicity, valence, and engagement. [2]Besides, it also detects user head position, gender, age, glasses, facial expressions, and eye tracking.

example, an expression of pain, in a situation in which eyes and/or brows cannot be seen, can be detected as a smile by these technologies (because of the stretching, open mouth).

As far as time is concerned, *Emotion API* and *Affectiva* show similar times to scan an image, while *Kairos* takes much longer to produce a result. Besides, the amount of values returned by *Affectiva* provides much more information to the developers, and it is easier to interpret the emotion that the user is showing than when we just have the weight of six emotions, for example. It is also remarkable the availability of *Affectiva*, which provides free services to those dedicated to research and education or producing less than $1,000,000 yearly.

2.2.3. Emotion from Text. There are certain situations in which the communication between two people, or between a person and a machine, does not have the visual component inherent to face-to-face communication. In a world dominated by telecommunications, words are powerful allies to discover how a person may be feeling. Although emotion detection from text (also referred as sentiment analysis) must face more obstacles than the previous technologies (spelling errors, languages, and slang), it is another source of affective information to be considered. Since emotion detection from texts analyzes the words contained on a message, the process to analyze a text takes some more steps than the analysis of a face or a voice. There is still a model that needs to be trained, but now text must be processed in order to use it to train a model [23]. This processing involves tasks of tokenization, parsing and part-of-speech tagging, lemmatization, and stemming, among others. Four technologies of this category are *Tone Analyzer* [24], *Receptiviti* [25], *BiText* [26], and *Synesketch* [27].

Due to the big presence of social media and writing communication in the current society, this field is, along with emotion detection from facial expressions, one of the most attractive fields to companies: posts from social media, messages sent to "Complaints" section, and so on. Companies which can know how their customers are feeling have an advantage over companies which cannot. Table 3 shows a comparative study of some of the key aspects of each technology. It is remarkable that as far as text is concerned, most of the companies offer a demo or trial version on their websites, while companies working on face or voice recognition are less transparent in this aspect. Regarding their accuracy, the four technologies have yielded good values. On the one hand, *BiText* has proved to be the simplest one, as it only informs if the emotion detected is good or bad. This way, the error threshold is wider and provides less wrong results. On the other hand, *Tone Analyzer* has proved to be less clear on its conclusions when the text does not contain some specific key words.

As far as the completeness of results is concerned, *Receptiviti* has been the one giving more information, revealing not only affective information but also personality-related information. The main drawback is that all these technologies (except *Synesketch*) are pay services and may not be accessible to everyone. Since *Synesketch* is not as powerful as the rest, it will require an extra effort to be used.

2.2.4. Emotion from Body Gestures and Movement. Even though people do not use body gestures and movement to communicate information in an active way, their body is constantly conveying affective information: tapping with the foot, crossing the arms, tilting the head, changing our position a lot of times while seated, and so on. Body language reveals what a person is feeling in the same way our voice does.

However, this field is quite new, and there is not a clear understanding about how to create systems able to detect emotions relating to body language. Most researchers have

Table 3: Comparison of emotion detection technologies from text.

Name	API/SDK	Requires Internet	Information returned	Difficulty of use	Free software
Tone analyzer	API	Yes	Emotional, social, and language tone	Low	No
Receptiviti	API	Yes	See [29]	Low	No
BiText	API	Yes	Valence (positive/negative)	Low	No
Synesketch	SDK	No	Six basic emotions	Medium	Yes

focused on facial expressions (over 95 per cent of the studies carried out on emotions detection have used faces as stimuli), almost ignoring the rest of channels through which people reveal affective information [28].

Despite the newness of this field, there are several proposals focused on recognizing emotional states from body gestures, and these results are used for other purposes. Experimental psychology has already demonstrated how certain types of movements are related to specific emotions [29]. For example, people experimenting fear will turn their bodies away from the point which is causing that feeling; people experimenting happiness, surprise, or anger will turn their bodies towards the point causing that feeling.

Since there are no technologies available for emotion detection from body gestures, there is not any consensus about the data we need to detect emotions in this way. Usually, experiments on this kind of emotion detection use frameworks (as for instance, *SSI*) or technologies to detect the body of the user (as for instance, Kinect), so the researches are responsible for elaborating their own models and schemes for the emotion detection. These models are usually built around the joints of the body (hands, knees, neck, head, elbows, and so on) and the angle between the body parts that they interconnect [30], but in the end, it is up to the researchers.

2.2.5. Emotion from Physiological States. Physiologically speaking, emotions originate on the limbic system. Within this system, the amygdala generates emotional impulses which create the physiological reactions associated with emotions: electric activity on face muscles, electrodermal activity (also called galvanic skin response), pupil dilatation, breath and heart rate, blood pressure, brain electric activity, and so on. Emotions leave a trace on the body, and this can be measured with the right tools.

Nevertheless, the information coming directly from the body is harder to classify, at least with the category system used in other emotion detection technologies. When working with physiological signals, the best option is to adopt a classification system based on a dimensional approach [25]. An emotion is not just "happiness" or "sadness" anymore, but a state determined by various dimensions, like valence and arousal. It is because of this that the use of physiological signals is usually reserved for research and studies, for example, related to autism. There are no emotion detection services available for this kind of detection based on physiological states, although there are plenty of sensors to read these signals.

In a recent survey on mobile affective computing [31], authors make a thorough review of the current literature on affect recognition through smartphone modalities and show the current research trends towards mobile affective computing. Indeed, the special capacities of mobile devices open new research challenges in the field of affective computing that we aim to address in the mobile version of the system proposed.

Finally, we can also find available libraries to be used in different IDEs (integrated development environments) supporting different programming languages. For instance, NLTK, in python [32] can be used to analyze natural language for sentiment analysis. Scikit-learn [33], also in python, provides efficient tools for data mining and data analysis with machine learning techniques. Lastly, OpenCV (Open Source Computer Vision Library) [34] supports C++, Python, and Java interfaces in most operating systems. It is designed for computer vision and allows the detection of elements caught by the camera in real time to analyze the facial points detected according, for instance, to the Facial Action Coding System (FACS) proposed by Ekman and Rosenberg [35]. The data gathered could be subsequently processed with the scikit-learn tool.

```
aff_information = get_affective_information()
#aff_information = {“face”: [...], “voice”: [...], “mimic”: [...]}
stress_flags = {“face”: 0.0, “voice”: 0.0, “mimic”: 0.0}
#values from 0 to 1 indicating stress levels detected
for er_channel, measures in aff_information:
    measure_stress(er_channel, measures, stress_flags)
if (stress_flags[“face”] > 0.6 and
  stress_flags[“voice”] < 0.3 and
  stress_flags[“mimic”] < 0.1):
  #reaction to affective state A
if (stress_flags[“face”] < 0.1 and
  stress_flags[“voice”] < 0.1 and
  stress_flags[“mimic”] < 0.5):
  #reaction to affective state B
...
```

3. Modifying the Behaviour of an Educational Software Application Based on Emotion Recognition

Human interaction is, by definition, multimodal [36]. Unless the communication is done through phone or text, people can see the face of the people they are talking to, listen to their voices, see their body, and so on. Humans are, at this point, the best emotion detectors as we combine information from several channels to estimate a result. This is how multimodal systems work.

It is important to remark that a multimodal system is not just a system which takes, for example, affective information from the face and from the voice and calculates the average of each value. The hard part of implementing one of these systems is to combine the affective information correctly. For example, a multimodal system combining text and facial expressions that detects a serious face and the message “it is very funny” will return “sarcasm/lack of interest,” while the result of combining these results in an incorrect way will return “happy/neutral.” It is proven that by combining information from several channels, the accuracy of the classification improves significantly.

For example, let us imagine we need to assess the stress levels of a person considering the affective information gathered through three different channels: affective information extracted from facial expressions, voice, and body language. Since we have more than one channel, we can support each measure taken from each channel with values detected in the others.

This way, it is possible not only to confirm with a high level of certainty the occurrence of an affective state, but also to detect situations that could not be sensed without performing multimodal emotion detection, as sarcasm.

The following code snippet shows an easy example of affective information combination. The mere fact of considering a measure in the context of more affective information gives us a whole new dimension of information.

To this end, we have developed an initial prototype in order to study how using multimodal emotion detection systems on educational software applications could enhance the user experience and performance. The proposed prototype, named emoCook, has been developed as a game to teach English to 9–11-year-old children. Information about this prototype can be found at [37]. At present, the prototype is only available in Spanish as it is initially addressed to Spanish-speaking children in the process of learning English.

The architecture of this application is shown in Figure 1. During the gameplay, the user is transmitting affective information (Figure 1-1) through their face, their voice, their behaviour, and so on. The prototype is receiving this information (Figure 1-3) and sending it to several third-party emotion detection services (Figure 1-4). After retrieving this information (Figure 1-5), we put it in context to extract conclusions from it about the user’s performance (Figure 1-6). Based on these results, the pace and difficulty level of the game changes (Figure 1-7), adapting it to the user’s affective state (Figure 1-2).

The theme of the game was focused on cooking issues to practice vocabulary and expressions related to this topic. It is organized in different recipes, from the easiest to hardest. Each recipe is an independent level and is divided into two parts. The first part is a platform game in which the player must gather all the ingredients needed to cook the recipe (Figure 2). The ingredients are falling from the sky all the time, along with other food we do not need for the recipe. If the player catches any food that is not in the ingredients list, it is considered as a mistake. The maximum number of mistakes allowed per level is five.

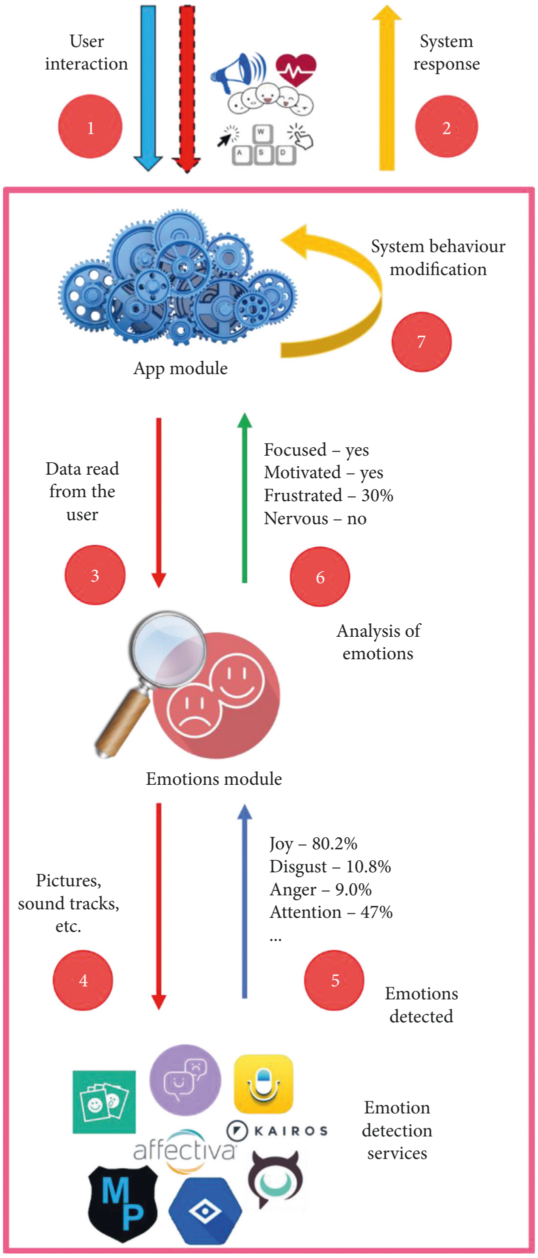


Figure 1: Application architecture.

After finishing this first part, the system shows a set of sentences (more or less complex) including vocabulary related to the recipe that the player has to read out loud to practice speaking and pronunciation. If the user fails thrice to read a sentence, the system will move to the next one, or finish the exercise if it is the last sentence.

This prototype has been implemented with three emotion detection technologies, which monitor the player’s

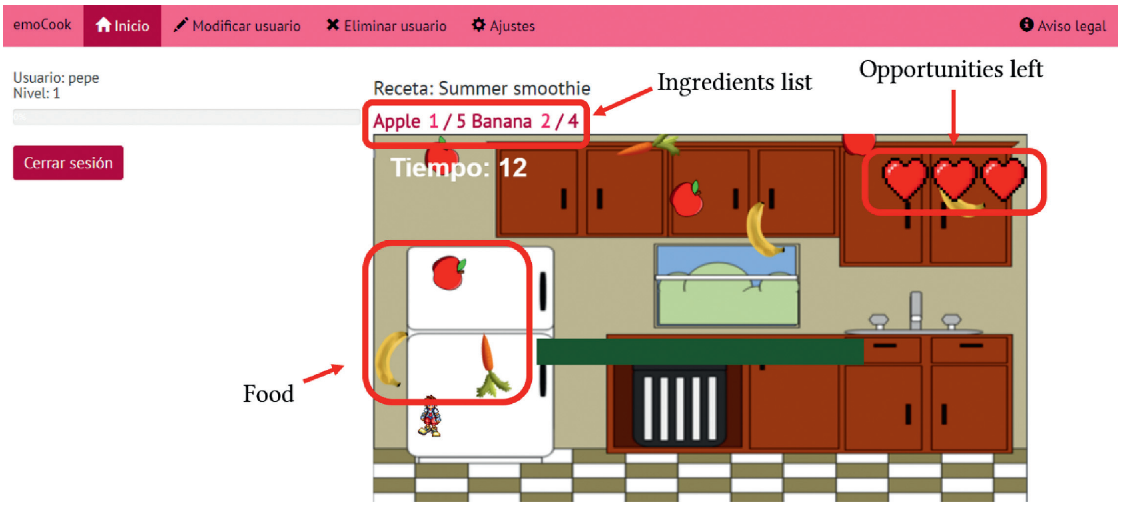


Figure 2: emoCook prototype.

affective state, and the results obtained are used to change the difficulty level and the pace of the game. Each time the player finishes a level, the affective data are analyzed, and according to the results, the difficulty of the next level is set. The technologies integrated in the system are the following:

(i) *Affectiva*. It uses the camera feed to read the facial expression of the player.

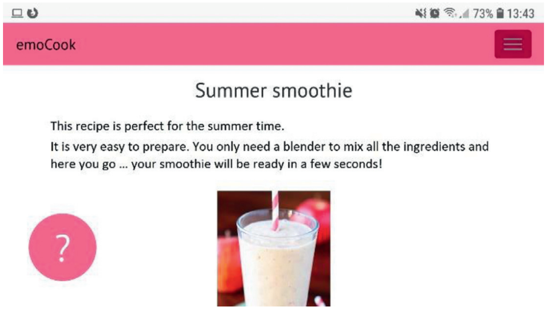
(ii) *Beyond Verbal*. It gathers the audio collected during the speech exercise to identify the affective state of the player attending to their speech features.

(iii) *Keylogger*. The game keeps a record of the keys pressed by the users, considering different factors: when they press a correct key, when they do not, when they press it too fast, and so on.

Because of changes on Beyond Verbal API, affective data from the speech could not be collected, so in the end, only data from the facial expression (using Affectiva) and from the behaviour when pressing keys (using Keylogger) were used. Affectiva is a third-party service, while Keylogger was developed within the prototype.

A mobile version of the system is also available, and it can be used through a browser running on a mobile device [37]. This way, the game can be controlled both with the arrow keys in a keyboard and by touching on a tactile screen. Touching on the left-hand side of the screen makes the character move to the left. Touching on the right-hand side of the screen makes the character move to the right and touching twice very quickly in any part of the screen makes the character jump upwards.

Figure 3 shows a screenshot of the mobile version of the application running in the Firefox browser in a mobile device. The possibility of using the system through a mobile device opens new ways of detecting emotions that we aim to explore in further research. For instance, we could use sensors such as the accelerometer or gyroscope to gather affective information. Initial trials have been performed with the API offered in [38] with promising results that will be further explored.



Summer smoothie

This recipe is perfect for the summer time.

It is very easy to prepare. You only need a blender to mix all the ingredients and here you go ... your smoothie will be ready in a few seconds!

Figure 3: Mobile version of emoCook system.

4. Evaluation of the System

In order to prove the initial hypothesis, the system has been assessed with real users by applying the method described in this section.

4.1. Participants and Context. We recruited sixteen children aged between 10 and 11 years old belonging to the same primary school and with a similar level of English knowledge to avoid differences in the education level that could affect the evaluation results. Their parents had been previously informed and authorised their participation in this evaluation. The setup of the experiment consisted of two laptops, one in front of the other so that participants could not see each other. Both laptops were equipped with mouse and webcam and Windows 10 as operating system and were connected to the same Wi-Fi network. The prototype was accessed through the browser Google Chrome in both laptops. This setup was prepared in a room the English teachers of the primary school provided us within the school premises.

4.2. Evaluation Metrics. The system was measured considering three types of metrics: effectiveness, efficiency, and satisfaction, that is, the users' subjective reactions when using the system. Effectiveness was measured by considering task completion percentage, error frequency, and frequency of assistance offered to the child. Efficiency was measured by calculating the time needed to complete an activity, specifically, the mean time taken to achieve the activity. Besides, some other aspects were also considered such as the number of attempts needed to successfully complete a level, number of keystrokes, and the number of times a key was pressed too fast as an indicative signal of nervousness.

Finally, satisfaction was measured with the System Usability Scale (SUS) slightly adapted for teenagers and kids [39]. This questionnaire is composed of ten items related to the system usage. The users had to indicate the degree of agreement or disagreement on a 5-point scale.

4.3. Experimental Design. After several considerations regarding the evaluation process for games used in learning environments [40], the following features were established:

(i) *Research Design.* The sample of participants was divided into two groups of the same size, being one of them the control group. This control group tested the application implemented without emotion detection and hence without modifying the behaviour of the application in real time according to the child's emotions. This one was called the System 2 group. The other group tested the prototype implemented with emotion detection which adapted its behaviour, by modifying the pace of the game and difficulty level, according to the emotions detected on the user, in such a way that if the user becomes bored, the system increases the pace of the game and difficulty level and on the contrary, if the user becomes stressed or nervous, the system decreases the speed of the game and difficulty level. This one was called the System 1 group. By doing this, it can be shown how using emotion detection to dynamically vary the difficulty level of an educational software application influences the performance and user experience of the students.

(ii) *Intervention.* The test was conducted in the premises of the primary school in a quiet room where just the participants (two at a time) using System 1 and System 2 and the evaluators were present. We prepared two laptops of similar characteristics, one of them running System 1 with the version of the application implemented with emotion recognition and the other laptop running System 2 with the version of the application without emotion recognition.

The whole evaluation process was divided into two parts:

(i) *Introduction to the Test.* At the beginning of the evaluation, the procedure was explained to the sixteen children at a time, and the game instructions for the different levels were given.

(ii) *Performing the Test.* Kids were called in pairs to the room where the laptops running System 1 and System 2 were prepared. None of the children knew what system they were going to play with. At the end of the evaluation sessions, the sixteen children completed the SUS questionnaire. Researchers were present all the time, ready to assist the participants and clarify doubts when necessary. When a participant finished the test, they returned to their classroom and called the next child to go in the evaluation room.

To keep the results of each participant fully independent, the sixteen users were introduced on the database of the prototype with the key "evalX," being "X" a number. Users with an odd "X" used System 1, while those with an even "X" were assigned to System 2 (control group).

The task that the participants had to perform was to play the seven levels of the prototype, including each level a platform game and a reading out loud exercise. The data collected during the evaluation sessions were subsequently analyzsed, and the outcomes are described next.

4.4. Evaluation Outcomes and Discussion. Although participants with System 1 needed, on average, a bit more time per level to finish (76.18 seconds against 72.7), we could appreciate an improvement on the performance of the participants using System 1, as most of them made less than 5 mistakes on the last level, while only one of the control group users of System 2 had less than 5 mistakes.

Figure 4 shows the evolution of the average number of mistakes, which increases in the control group (System 2) from level 4 onwards. Since the game adapts its difficulty (in System 1), after detecting a peak of mistakes in the fourth level (as a sign of stress, detected as a combination of negative feelings found in the facial expression and the way the participant used the keyboard), the difficulty level was reduced. This adaptation made the next levels easier to play for participants using System 1, what was reflected in less mental effort. Since participants using System 2 did not have this feature, their average performance got worse.

On average, participants using System 1 needed 1.33 attempts to finish each level, while participants using System 2 needed 1.59, almost 60% more. Also, the ratio of mistakes to total keystrokes was also higher in the case of System 2 users (19% against the 12% from users of System 1). Likewise, System 2 users asked for help more often (13 times) than System 1 users (10 times). In future experimental activities, the sample size would be increased in order to obtain more valuable data.

The evaluation was carried out as a between-subjects design with *emotion recognition* as the independent variable (using or not using emotion recognition features) and

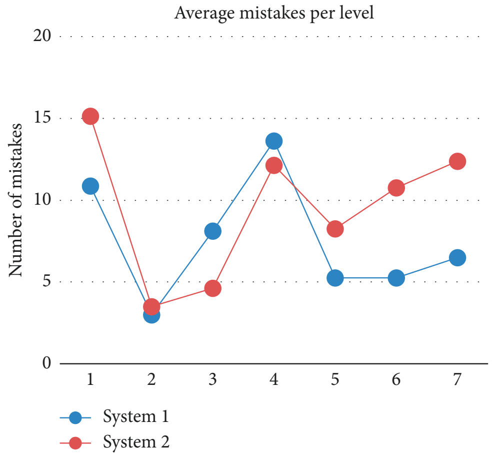


Figure 4: Average mistakes per level.

attempts (attempts needed to finish each level), *time* (time (seconds) needed to finish each level), *mistakes* (number of mistakes), *keystrokes* (number of keystrokes), and *stress* (number of times a key was pressed too fast in a short time) as the dependent variables.

We performed a standard t-test [41] to compare the means of each dataset and test the null hypothesis that there was no significant difference in the students' performance when using emotion recognition to adapt the system behaviour. We used $\alpha = 0.05$ as our limit for statistical significance, with significant results reported below.

Regarding *keystrokes* ($t = 0.97$; $p = 0.666$), *mistakes* ($t = -1.51$; $p = 0.26$), and *stress* ($t = 1.13$; $p = 0.51$), t-test results confirmed the null hypothesis was false and, thus, that the two datasets are significantly different.

Although the dependent variables *time* ($t = 0.44$; $t = 1.31$) and *attempts* ($t = -0.42$; $t = 1.33$) were similar in both datasets, the efficiency (considered as the lowest number of actions a user needs to finish each level) is greater in users of System 1, even though both users of System 1 and System 2 finished within a similar time frame, what helped the first ones to make less mistakes. The outcomes of the evaluation shown in Figure 4 indicate a clear improvement when using System 1 as the number of mistakes increases in users of System 2 at higher difficulty levels.

Finally, Table 4 and Table 5 show the results of the SUS scores per system and participant. The final value is between 0 and 100, 100 being the highest degree of user's satisfaction. As we can see, System 1 users rated the application with a higher level of satisfaction compared to the level obtained by users of System 2, as shown in Figure 5.

Table 4: Satisfaction results for System 1.

Participant	SUS score
1	90
3	90
5	90
7	100
9	92.5
11	87.5
13	82.5
15	80
Mean	89.06

Table 5: Satisfaction results for System 2.

Participant	SUS score
2	92.5
4	75
6	87.5
8	85
10	75
12	92.5
14	72.5
16	90
Mean	83.75

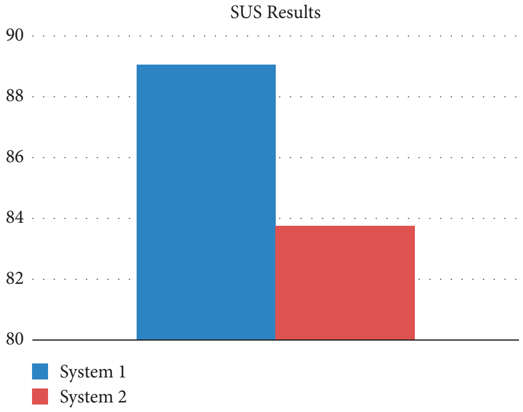


Figure 5: Comparison of SUS results in both systems.

5. Conclusions and Final Remarks

Emotion detection, together with Affective Computing, is a thriving research field. Few years ago, this discipline did not even exist, and now there are hundreds of companies working exclusively on it, and researchers are investing time and resources on building affective applications. However, emotion detection has still many aspects to improve in the coming years.

Applications which obtain information from the voice need to be able to work in noisy environments, to detect subtle changes, maybe even to recognize words and more complex aspects of human speech, like sarcasm.

The same applies for applications that detect information from the face. Most people use glasses nowadays, which can greatly complicate accurate detection of facial expressions.

Applications able to read body gestures do not even exist now, even though it is a source of affective information as valid as the face. There are already applications for body detection (Kinect), but there is no technology like *Affectiva* or *Beyond Verbal* for the body yet.

Physiological signals are even less developed, because of the imposition of sensors that this kind of detection requires.

However, some researchers are working on this issue so physiological signals can be used as the face or the voice. In a not too distant future, reading the heartbeat of a person with just a mobile with Bluetooth may not be as crazy as it may sound.

Previous technologies analyze the impact of an emotion in our bodies, but what about our behaviour? A stressed person usually tends to make more mistakes. In the case of a person interacting with a system, this will be translated in faster movements through the user interface, or more mistakes when selecting elements or typing, and so on. This can be logged and used as another indicator of the affective state of a person.

All these technologies are not perfect. Humans can see each other and estimate how other people are feeling within milliseconds, and with a small threshold error, but these technologies can only try to figure out how a person is feeling according to some input data. To get more accurate results, more than one input is required, so multimodal systems are the best way to guarantee results with the highest levels of accuracy.

In this paper, we present an educational software application that incorporates affective computing by detecting the users' emotional states to adapt its behaviour to the emotions detected. Assessing this application in comparison with another version without emotion detection, we can conclude that the user experience and performance is higher when including a multimodal emotion detection system. Since the system is continuously adapting itself to the user according to the emotions detected, the level of difficulty adjusts much better to their real needs.

On the basis of the outcomes of this research, new challenges and possibilities in other kind of applications will be explored; for example, we could "stress" a user in a game if the emotions detected show that the user is bored. The application could even introduce dynamically other elements to engage the user in the game. What is too simple bores a user, whereas what is too complex causes anxiety. Changing the behaviour of an application dynamically according to the user's emotions, and also according to the nature of the application, increases the satisfaction of the user and helps them decrease the number of mistakes.

As future work, among other things, we aim to improve the mobile aspects of the system and explore further the challenges that the sensors offered by mobile devices bring about regarding emotion recognition, especially in educational settings.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research work has been partially funded by the regional project of JCCM with reference SBPLY/17/180501/000495, by the Scholarship Program granted by the Spanish Ministry of Education, Culture and Sport, and the predoctoral fellowship with reference 2017-BCL-6528, granted by the University of Castilla-La Mancha. We would also like to thank the teachers and pupils from the primary school "Escolapios" who collaborated in the assessment of the system.

References

[1] R. W. Picard, *Affective Computing*, MIT Press, Cambridge, UK, 1997.

[2] E. Johnson, R. Hervás, C. Gutiérrez, T. Mondéjar, and J. Bravo, "Analyzing and predicting empathy in neurotypical and nonneurotypical users with an affective avatar," *Mobile Information Systems*, vol. 2017, Article ID 7932529, 11 pages, 2017.

[3] S. Koelstra, C. Muhl, M. Soleymani et al., "DEAP: a database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[4] R. Kaliouby, "We need computers with empathy," *Technology Review*, vol. 120, no. 6, p. 8, 2017.

[5] S. L. Marie-Sainte, M. S. Alrazgan, F. Bousbahi, S. Ghouzali, and A. W. Abdul, "From mobile to wearable system: a wearable RFID system to enhance teaching and learning conditions," *Mobile Information Systems*, vol. 2016, Article ID 8364909, 10 pages, 2016.

[6] M. Li, Y. Xiang, B. Zhang, and Z. Huang, "A sentiment delivering estimate scheme based on trust chain in mobile social network," *Mobile Information Systems*, vol. 2015, Article ID 745095, 20 pages, 2015.

[7] B. Ovcjak, M. Hericko, and G. Polancic, "How do emotions impact mobile services acceptance? A systematic literature review," *Mobile Information Systems*, vol. 2016, Article ID 8253036, 18 pages, 2016.

[8] P. Williams, "Emotions and consumer behavior," *Journal of Consumer Research*, vol. 40, no. 5, pp. viii–xi, 2014.

[9] E. Andrade and D. Ariely, "The enduring impact of transient emotions on decision making," *Organizational Behavior and Human Decision Processes*, vol. 109, no. 1, pp. 1–8, 2009.

[10] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 55–64, 2003.

[11] R. W. Picard, "Affective computing," Tech. Rep. 321, M.I.T Media Laboratory Perceptual, Computing Section, Cambridge, UK, 1995.

[12] I. Morgun, *Types of Machine Learning Algorithms*, 2015.

[13] J. García-García, V. Penichet, and M. Lozano, "Emotion detection: a technology review," in *Proceedings of XVIII International Conference on Human Computer Interaction*, Cancún, México, September 2017.

[14] S. Casale, A. Russo, G. Scebba, and S. Serrano, "Speech emotion classification using machine learning algorithms," in *Proceedings of IEEE International Conference on Semantic Computing 2008*, pp. 158–165, Santa Monica, CA, USA, August 2008.

[15] Beyond Verbal, "Beyond verbal–the emotions analytics," May 2017, http://www.beyondverbal.com/.

[16] Vokaturi, May 2017, https://vokaturi.com/.

[17] T. Vogt, E. André, and N. Bee, "EmoVoice—a framework for online recognition of emotions from voice," in *Perception in Multimodal Dialogue Systems*, E. André, L. Dybkjær, W. Minker, H. Neumann, R. Pieraccini, and M. Weber, Eds., Springer, Berlin, Heidelberg, Germany, 2008.

[18] Good Vibrations, "Good vibrations company B.V.–recognize emotions directly from the voice," May 2017, http://good-vibrations.nl.

[19] Microsoft, "Microsoft cognitive services–emotion API," May 2017, https://www.microsoft.com/cognitive-services/en-us/emotion-api.

[20] Affectiva, "Affectiva," May 2017, http://www.affectiva.com.

[21] nViso, "Artificial intelligence emotion recognition software-nViso," May 2017, http://nviso.ch/.

[22] Kairos, "Face recognition, emotion analysis & demographics," May 2017, https://www.kairos.com/.

[23] H. Binali and V. Potdar, "Emotion detection state of the art," in *Proceedings of the CUBE International Information Technology Conference on-CUBE '12*, pp. 501–507, New York, NY, USA, September 2012.

[24] IBM, "Tone analyzer," May 2017, https://tone-analyzer-demo.mybluemix.net/.

[25] Receptiviti, May 2017, http://www.receptiviti.ai/.

[26] Bitext, "Bitext API," May 2017, https://api.bitext.com.

[27] U. Krčadinac, "Synesketch: free open-source textual emotion recognition and visualization," May 2017, http://krcadinac.com/synesketch/.

[28] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 15–33, 2013.

[29] R. W. Picard, "Future affective technology for autism and emotion communication," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3575–3584, 2009.

[30] Universität Augsburg University, "OpenSSI," May 2017, https://hcm-lab.de/projects/ssi/.

[31] E. Politou, E. Alepis, and C. Patsakis, "A survey on affective computing," *Computer Science Review*, vol. 25, pp. 79–100, 2017.

[32] N. Hardeniya, *NLTK Essentials*, Packt Publishing Limited, Birmingham, UK, 2015.

[33] G. Hackeling, *Mastering Machine Learning with Scikit-Learn*, Packt Publishing Limited, Birmingham, UK, 2014.

[34] J. Howse, P. Joshi, and M. Beyeler, *OpenCV: Computer Vision Projects with Python*, Packt Publiser Limited, Birmingham, UK, 2016.

[35] P. Ekman and E. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, Oxford University Press, Oxford, UK, 2005.

[36] J. Tao and T. Tan, "Affective computing: a review," *Lecture Notes in Computer Science*, vol. 3784, pp. 981–995, 2005.

[37] J. M. Garcia-Garcia, "emoCook," December 2017, https://emocook.herokuapp.com/.

[38] A. Bar, "What web can do today. An overview of the device integration HTML5 APIs," June 2018, https://whatwebcando.today/device-motion.html.

[39] Usability.gov, "Usability.gov-improving the user experience," January 2018, https://www.usability.gov/get-involved/blog/2015/02/working-with-kids-and-teens.html.

[40] A. All, E. P. Nuñez Castellar, and J. Van Looy, "Assessing the effectiveness of digital game-based learning: best practices," *Computers & Education*, vol. 92-93, pp. 90–103, 2016.

[41] D. Garson, *Significance Testing: Parametric and Nonparametric*, Statistical Associates Publishing, Blue Book Series, Asheboro, NC, USA, 2012.

3.2 Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions

Citación: Garcia-Garcia, J. M., Penichet, V. M. R., Lozano, M. D., and Fernando, A. (2021). Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions. Universal Access in the Information Society, ISSN: 1615-5297. Volumen 21, páginas 809–825. https://doi.org/10.1007/s10209-021-00818-y

Índice de impacto: 2.629, Q3 (JCR 2021)

Esta publicación avala esta tesis por compendio de publicaciones.

Universal Access in the Information Society
https://doi.org/10.1007/s10209-021-00818-y

LONG PAPER

Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions

Jose Maria Garcia-Garcia[1] · Victor M. R. Penichet[1] · Maria D. Lozano[1] · Anil Fernando[2]

Accepted: 26 April 2021

Abstract

Autism spectrum disorder (ASD), which since 2013 is considered as an umbrella term for several disorders such as autistic syndrome, Asperger's disorder and pervasive developmental disorder, is characterized, among other aspects, by deficits in social-emotion reciprocity. This deficit manifests itself as a reduced sharing of emotions and an increased difficulty in interpreting emotions other people are feeling, which in the end leads to more impairments in social communication. Since it is possible to help a person with ASD (especially children) to improve their ability to understand and detect emotions, we have developed a proposal which integrates emotion recognition technologies, often used in the field of HCI, to try to overcome this difficulty. In this paper, we present a novel software application developed as a serious game to teach children with autism spectrum disorder (ASD) to identify and express emotions. The system incorporates cutting-edge technology to support novel interaction mechanisms based on tangible user interfaces (TUIs) and emotion recognition from facial expressions. In this way, children interact with the system in a natural way by simply grasping objects with their hands and using their faces. The system has been assessed on the premises of an association with children with ASD. The outcomes of the evaluation are very positive and support the validity of the proposal.

Keywords HCI · Affective computing · Emotion recognition · ASD

1 Introduction

Emotion detection has recently become an important research topic. In the last ten years, up to 300.000 papers about emotion detection have been published, according to Google Scholar [22]. Although a part of this research effort is focused on *creating* emotion detectors, there is also a big effort dedicated to the integration of these detectors into final products in order to improve the user experience. Being able to know how users feel while using a product and, more importantly, being able to change the product's behavior so that the user experience is the best possible for each specific user, is a powerful tool that was not previously available in the field of human–computer interaction research. Moreover, this information about a person's emotions is valuable not just for the researchers studying a product's user experience or users' behavior, but for the users themselves. Emotional awareness, or the ability to be aware of, and identify, internal emotional states [47] leads us to having a better understanding of our own emotions and to being able to better regulate the affect within ourselves and others, which contributes to improving our well-being [45]. Developing our emotional awareness also helps us in building our emotional intelligence, which has been proved to benefit individuals in several dimensions of their lives, including their academic and professional life [34].

Emotional awareness, i.e., the ability to recognize one's own emotions, and emotional intelligence in general, is an ability that humans learn and develop throughout their lives. However, not every individual is equally able to cultivate this skill. For instance, people with autism spectrum

✉ Maria D. Lozano
maria.lozano@uclm.es

Jose Maria Garcia-Garcia
josemaria.garcia@uclm.es

Victor M. R. Penichet
victor.penichet@uclm.es

Anil Fernando
w.fernando@surrey.ac.uk

[1] Albacete Research Institute of Informatics, Universidad de Castilla-La Mancha, Albacete, Spain

[2] Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK

Published online: 19 June 2021

disorder (ASD) find it hard to recognize emotions in others, as well as to understand and handle their own emotions [4]. This impairment (external emotion recognition) is related to their problems with paying attention and distinguishing faces. Fortunately, these emotional-intelligence-related skills can be learnt and trained, as we can see in the existent literature. In [12], Dawson et al. reviewed studies covering a long period of time which show how children with ASD can overcome one of the challenges that people with this disorder confront: facial recognition. Daou et al. [11] also reviewed existing literature to collect studies about teaching emotion expression and recognition to children with ASD, and most of these studies reported positive results. Since ASD is mostly correctly diagnosed during early childhood, the sooner this emotional intelligence education starts, the easier it will be for children to apply this knowledge in adulthood [12]. Technology has proven to be a powerful ally in this learning process. Yeni et al. [56] reviewed studies which used educational mobile applications (that is, applications running on mobiles or tablets) to teach different abilities to people with intellectual disabilities, showing that not only do they accept technology very easily, but also that they can become fluent using the portable device on which the application is running and learn the skill they are working on with constant practice. Nisiforou et al. [36] took a step forward in this type of literature reviews by examining works that use technology to teach skills specifically to children with disabilities, showing the popularity of works involving games or educational applications and robots.

Games are, in fact, an effective tool to teach new things to children [37]. Nowadays, when games are used with a purpose that is not simply entertainment they are referred to as *serious games*. Serious games aim to promote learning through entertainment, exploiting the cognitive benefits of games to ease the learning process [14]. Games can catch children's attention, and this also includes children with ASD [24]. However, a game must meet a certain criterion to keep this attention, i.e., not every game is equally appealing for every person. Even when a serious game is used in a traditionally serious environment (e.g., a classroom), it must keep its fun component and it must keep the players engaged with it, introducing variability and challenges, ensuring that it is difficult enough to prevent the players from getting bored, but easy enough to avoid frustration. According to the existing literature, serious games have previously been used to teach emotional-intelligence-related skills with positive results [33], and we have based our proposal on this knowledge.

Considering the context described above, we have developed a serious game for Android devices combining affective computing and tangible user interfaces (TUI) [30]. Emotion detection technology is used to detect what emotion the player is expressing, in order to help them develop their emotional awareness, and tangible user interfaces are used because of their well-known beneficial influence on learning [30][43], and to overcome the difficulties that children with ASD experience when using computers.

The game consists of three parts, each one with its own goal and involving a different interaction style in order to use this variability to attract the player's attention. These different parts are, in turn, three different games:

- *Game 1—Recognition of emotions using TUI.* This game starts with players being prompted with a picture depicting an emotion in the app. While keeping this emotion in mind, they must go through a set of physical cards, each of them used as tangible objects by using NFC technology, looking for the card which best represents the emotion prompted. After the player has decided which card best expresses this emotion, they must bring it close to the NFC reader, i.e., the device where the app is running. If the card chosen is the right one, the player is prompted with the next emotion. If the chosen card does not represent the requested emotion, an error message is shown;
- *Game 2—Depiction of emotion.* In the second game, the players are prompted again with a picture depicting an emotion. However, in this case they must express this emotion themselves. Using the device's camera, emotion detection services are used to recognize what emotion the player is expressing with his/her face;
- *Game 3—Recognition of emotions in the wild using TUI.* In this last phase, players have to recognize emotions again. However, instead of being prompted with a picture, they are shown a piece of video in which a specific emotion is being displayed. Using the cards from the first game (tangible user interfaces), the players must indicate what emotion is being shown in the video they are watching.

As was specified by the user requirements, in order to represent emotions in games 1 and 3 we have used pictograms as well as pictures of real people, so players have different references to learn about feelings and how to identify them. We turned these pictures into tangible user interfaces by using NFC tags hidden within them. Previously, these tags were programmed with their corresponding emotion name, e.g., the NFC tag attached to the picture of a woman smiling had the value "Happy" loaded onto it. For the part of the game entailing emotion detection, after reviewing several technologies [19] we chose *Affectiva*, which offers facial-expression-based emotion detection via a straightforward SDK, and was easy to integrate into the main app. As part of the application usage flow, players have to log in using their credentials (previously, they should have signed up in the app) and then choose the game they want to play.

The target audience of the application is children aged between 6 and 12 years old with ASD, so it is an audience mature enough to use electronic devices safely but still at an early phase in their education, so the learning of new emotional skills has a greater impact.

Regarding the evaluation of this tool, we carried out a preliminary evaluation of the system with the assistance of experts in the problem domain from the Association "*Autism Development*" to assess both the teaching capabilities of the tool and its usability. Two specialists from the Association, as well as three children with ASD, participated in this assessment, whose goal was to assess the acceptance of the system as a suitable tool to be used to teach emotions to children with ASD and to test the usability of the first prototype of this application.

It is important to point out that the key contribution of this article lies in the integration of *automatic emotion recognition technologies* and *tangible user interfaces* in a serious game in order to obtain a more natural interaction, which is an essential feature especially when working with children with ASD. While the use of games and software applications to teach skills to children with ASD is not something new, this kind of tools usually requires the assistance of a therapist, a parent or a caretaker. By using automatic emotion recognition and a more natural interaction mechanism (TUI), we have achieved a very user-friendly application, as the evaluation with experts has shown.

This paper is divided into six sections, including the current one. Section 2 introduces some key concepts that support the decisions taken during the development of the system. We provide a detailed description of the system developed in Sect. 3. Section 4 presents the evaluation process carried out with children with ASD and specialists from the Association. Section 5 reviews the outcomes of the evaluation process. Finally, Sect. 6 presents the main conclusions and lines for future work.

2 Background concepts and related work

2.1 Autism spectrum disorder

According to the Diagnostic and Statistical Manual of Mental Disorders, autism spectrum disorder (ASD), commonly referred to as autism, is a neurodevelopmental disorder characterized by persistent deficits in social communication and social interaction across multiple contexts and restricted, repetitive patterns of behavior, interests, or activities, with these symptoms being shown in the early developmental period [4]. Every case of autism is unique since autism encompasses a whole spectrum: some cases may be mild, while others may be severe with regard to the symptoms. In fact, in 2013 the term ASD became in an umbrella term for a set of behavior disorders, namely early infantile autism, childhood autism, Kanner's autism, high-functioning autism, atypical autism, pervasive developmental disorder, childhood disintegrative disorder, and Asperger's disorder. Furthermore, since there is no cure [7], early diagnosis is very important, since the sooner this disorder is detected, the sooner the treatment can begin. Treatment includes occupational therapy, applied behavioral analysis, sensory integration therapy, etc. [51]. Again, although autism is not a curable disorder, the aforementioned treatments can help decrease the social deficits associated with ASD.

As part of the diagnosis process, the person must be assessed to establish the presence and/or severity of the symptoms. For instance, in the case of impairments related to communication and social interaction, these symptoms are pervasive in all kinds of social interaction and sustained in time. In order to obtain the most reliable and valid assessment of these impairments, we must gather all the information available: clinicians' observations, the caregiver's history, colleagues' impressions and, when possible, a self-report. Not only do we need to assess impairments in communication, but also deficits in social-emotional reciprocity. Even though people with ASD may be able to communicate correctly from a formal point of view (correct grammar, good vocabulary, etc.), they may still struggle to engage in social interaction due to not knowing what tone or attitude adopt on each occasion, not understanding how the other person is feeling, avoiding eye contact, etc. Lack of reciprocity is what characterizes social interaction with a person suffering from ASD [4].

Another characteristic symptom of ASD is restricted, repetitive patterns of behavior, activity or movement. Examples of this repetition are repeating movements over and over, aligning or ordering objects in a specific way, the parroting of heard words (echolalia), etc. This repetition also manifests itself through the adoption of routines, the ritualization of certain patterns (doing something by always following the same sequence of tasks), something that in the end evolves into a huge resistance to change. These routines can sometimes be the result of hypo- or hyperreactivity to certain stimuli, that is, an excessive fascination for, or rejection of, something involving taste, smell, texture or appearance, or rituals involving these senses.

In short, poor social skills and emotional instability are inherent in people with ASD, with the severity of these deficits varying to a great extent depending upon the type of ASD the person has. Even though autism (ASD) does not have, strictly speaking, a cure, therapy can help people with this disorder become more independent and improve their social communication and interaction skills. One of the methods applied in therapy is *social skill groups*. This type of therapy takes place once a week over 12 weeks or more and entails a group of between two and six individuals with ASD being led by one to three therapists. During these sessions, which last from 60 to 90 min, the therapists give a lesson about a specific social skill, including role playing, to practice this skill and promote a discussion about the whole lesson [44]. This form of therapy affects a person's social functioning by providing a learning environment that allows immediate rehearsal and practice.

In contrast to this form of therapy, other types of therapy focus on improving communication in general, and for this purpose different protocols are deployed. For instance, when the goal of a therapy is speech production, speech imitation protocols are used. However, this approach presents some drawbacks. For example, since the subjects learn to imitate the teacher's speech in a formal environment, they fail to generalize this new skill to new environments and social interactions [7].

As alternatives to speech imitation, other communication protocols have been used, such as sign language or picture-based communication systems. Nevertheless, like speech imitation, these protocols present several drawbacks, such as the difficulty of learning sign language, or the inaccuracy of picture-point systems. These systems usually fail because they do not consider the point of view of the child. For instance, they assume that once the child knows the word to name something or someone, he or she will be able to use it in all contexts [8], and this is simply not true.

The Picture Exchange Communication System (PECS) proposed in 1994 [8] represented an approach that corrected the flaws of previous communication systems. This system suggests several phases, all of them with their own prompting, reinforcement, and error correction strategies, based on the principles of applied behavioral analysis, to teach spontaneous, functional communication to children. In the course of these phases, children are taught how to communicate using pictograms, to go through their pictograms to find the most suitable image for some answer, how to prompt a social interaction, how to comment on something, etc. [7].

Apart from PECS, other proposals based on showing pictures, particularly pictures of faces, to children with ASD have been made. In [21], Golan et al. developed a children's animation series called "The Transporters," which was about eight characters who were vehicles with human faces, to teach children with ASD about facial expressions and emotions expressed through this mechanism. According to the systemizing theory of autism, individuals with ASD have intact, or even enhanced, *systemizing skills*, which help them understand and analyze rule-based systems, find patterns, and so on. A good example of a rule-based system is vehicles such as trains or cable cars, which only move back and forth along linear tracks, making them a predictable system. This series was built upon the following idea: since the vehicles in the show, which have faces of several actors and actresses expressing emotions attached to them, make up a "safe space," children will pay more attention to them (even without realizing they are doing so) instead of avoiding the faces, helping them learn about emotional expressions [21]. Faces on vehicles are considered a "safe space" by children with ASD because vehicles are rule-based systems, they are predictable, in contrast to human bodies, which move in unexpected ways. This study has been replicated and reviewed [5]1155, and the results, as well as their generalization, appear to be valid. In our game, we take advantage of this concept, namely the idea of attaching faces to predictable elements, in the form of tangible user interfaces, though PECS, using pictograms together with real photographs, to teach the different emotions to children and to help them generalize this knowledge. In our setup, each physical image becomes a tangible user interface that children will use to tell the app what emotion they have been requested to recognize during the different games.

2.2 Emotions

While they are pervasive in every aspect of our lives and are receiving more and more attention every day, emotions are still difficult to define and classify. If we start tracing the definitions of emotions back in time, we find endless debates and endless definitions. Aristotle proposed his own taxonomy of emotions in 400 B.C. The catalogue of proposals is so extensive that, in 1981, the authors of [28] gathered 92 different definitions of emotion, each one considering a different aspect of the same topic. For now, and following the trends of affective computing over the last few years, we will take an emotion to be a physical reaction of the body, caused by the limbic system, to some event or circumstance. This reaction can be either perceptible for external observers (changes in the tone of voice, facial expressions, body gestures) or imperceptible (heartbeat, electrical brain activity, etc.). We will look at this more closely in the following subsection.

Two of the most popular proposals regarding emotions and their classification were made by Robert Plutchik and Paul Eckman. Robert Plutchik proposed a model based on a 2D/3D "flower" of emotions. In Plutchik's model, called the wheel of emotions (Fig. 1), every human emotion is a combination of several primary emotions, namely ecstasy, admiration, terror, amazement, grief, loathing, rage and vigilance [42]. Each primary emotion can lead to others, depending on the degree of intensity with which someone feels it. The rest of the emotions are combinations of these primary emotions.

The other proposal was made by the psychologist Paul Ekman. One of the topics Ekman studied was the universality of emotions. Following the Darwinian view of emotions, Ekman wanted to prove that emotions, or at least a subset of them, were inherent in every human [15]. As part of this

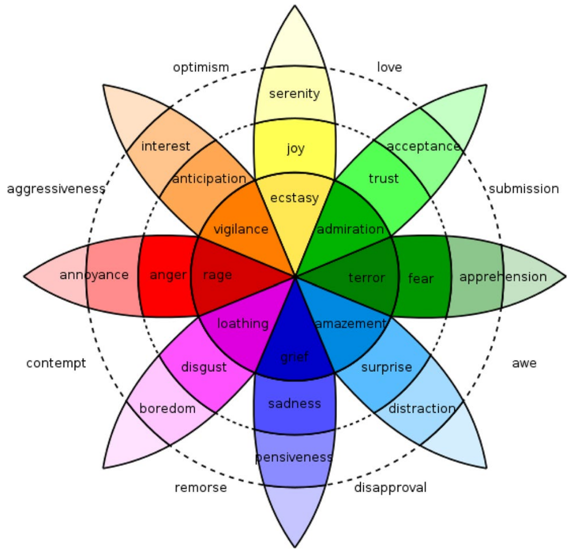


Fig. 1 Wheel of emotions [53]

study, he developed the Facial Action Codification System, a system which identifies 42 points on the face, the eyes and the head and uses them to identify an emotion [16]17. In this way, a facial expression can be defined by the position of a set of 42 points on the face. By codifying the human face as a set of numbers, FACS opened the door to the creation of emotion detectors based on facial expression using machine learning and automatic classifiers.

As per other studies carried out by Ekman, he discovered that there were six emotions which were universal to every human being, regardless of culture or education, since they were hardcoded into our DNA, following the Darwinian explanation for the origin of emotions [15]. These universal emotions are joy, sadness, anger, surprise, fear and disgust. However, since this study was published, some researchers have reviewed it, and find flaws and holes in this universality [52]. Some of these studies have even proposed a different number of universal emotions [26]. Despite the imperfections in Ekman's theory, the six-basic-emotions approach is widely extended in the affective computing field, it being the de facto classification system used by emotion detectors to express what emotions have been found.

With regard to children with ASD, in order to get them to express an emotion, they must first learn how to identify it, and this is what games or practice exercises are for. Therefore, the game we have developed has three well-distinguished parts: a first part in which the children learn to identify an emotion, a second part in which they learn how to express it themselves, and, finally, a third part in which they learn to identify these emotions in the wild, in a spontaneous situation.

2.3 Affective computing

Affective computing (AC), as it was defined in 1995, is any form of computing that relates to, arises from, or influences emotions [40]. Although affective computing presents several lines of work, one of the most popular is automatic emotion detection, i.e., the use of automatic classifiers to detect emotions in a voice, in a face, etc. As we mentioned in the previous subsection, in the field of automatic emotion detection, we understand an emotion to be its physical manifestation in the body. Hence, emotion detection encompasses the detection of those physical manifestations and the subsequent analysis of those signals.

In order to read these data, we need different types of sensors, depending on what type of information is going to be collected [19]. For instance, to read someone's facial expression or body language, we need a camera or some device such as Kinect that allows us to capture images and to track the human body. Along these lines, there are also devices that allow us to track the user's eyes, which, in the end, means that we can know what the user is looking at, for how long, etc. If we wish to analyze someone's voice, we need a microphone; to read things such as someone's heartbeat or muscle activity, we need more invasive tools, such as a wristband with sensors, electrodes placed on the part whose electrical activity we are going to measure, with all this connected to a controller, such as a computer or a Raspberry microcontroller.

Regarding emotion detection, a new trend has emerged in the field of HCI research that is based on the following idea: what if we could detect how users are feeling while using an application or system and use this information to change the behavior of that system in order to make the users' experience as good as possible? With this idea in mind, much research work has been carried out over the past few years, and indeed our work forms part of this new trend for combining HCI and AC. However, marketing and user experience (UX) are not the only fields which are applying these AC-related technologies. In [31] the authors review how affective computing technologies and strategies have been applied to improve the lives of children with ASD. For this purpose, computer software that detects behavioral signs of emotions and models emotional functioning was used. Facial expressions, vocalizations, electrodermal activity, affective intelligent tutoring systems have been used to try to help children with ASD overcome their social deficits, and these studies obtained acceptable results, although there is a consensus about the need to continue replicating these experiments in order to actually systematize the application of AC technologies in ASD-related therapies [31],35.

With regard to the affective computing component included in this work, we have focused on emotion detection based on facial expressions, and we have integrated this kind of detection technology in our software application as part of one of the games. We chose this form of emotion detection over the other ones after reviewing the different emotional channels from which we can read affective information from the users, the different tools needed to gather that information, the availability of these tools and, also, after considering our previous experience with other emotion detectors [20].

The decision to choose facial expression over any other emotional channel was taken on the basis of the "universality" of facial expressions to express emotions, and the relative maturity of this type of automatic emotion detection. In the following subsection, the technologies used in the application developed are described.

2.4 Related work

Prior to the development of the app, we reviewed several off-the-shelf applications of a similar nature. A list of applications developed for children with ASD to help them in their daily activities and to learn about emotions can be found in Table 1. By taking a quick look at the table, we can see that, in this sample, all the reviewed systems are used on a touch-based device, usually a tablet or a smartphone. It is also noteworthy that only one application, *Emotionalyser* uses facial recognition, and just a few of them use physical images as part of the game.

We also reviewed the existing literature regarding serious games applied in the education of children with ASD to find possible gaps in this research field. Reviews such as [33] and [57] reveal the state of the art regarding serious games applied to therapy for children with ASD. For instance, most current proposals are designed to be used on a desktop or a laptop computer. In [57], 40 papers are reviewed and 70 per cent employ serious games running on computers, disregarding more usual or natural interaction mechanisms such as touch screens or tangible user interfaces. With regard to automatic emotion recognition, the use of facial-expression-based emotion detectors is quite scarce, but it is present in some approaches using serious games [18, 25, 27, 32, 39, 46, 49]. It is worth highlighting that most of the reviewed games using emotion recognition use a "*homemade*" emotion detector instead of resorting to off-the-shelf solutions or open-source software, with the impact in time and cost that this fact has on the process of developing a game.

For the sake of completeness, we decided to also review other studies focused on teaching skills to children with ASD. Artoni et al. [3] developed a web application, namely ABCD SW, consisting of several types of exercises, in such a way that it can be used by the children while a tutor or a parent is monitoring their progress from another device. Studies such as [21] and [57] have used the animated series "The Transporters," which shows different vehicles such as trains and cable cars with human faces, to teach children with ASD to recognize faces and emotions. These studies

Table 1 Comparison of related works

Name	Platform (Android/iOS/Web/Windows)	Camera (Yes/No)	Facial recognition (Yes/No)	Real pictures (Yes/No)	Tactile interface (Yes/No)
Happy baby	Android	No	No	No	Yes
Emotionalyser	Android	Yes	Yes	Yes	Yes
Camp Discovery	Android/iOS	No	No	No	Yes
Touch-Emotions	Android/iOS	No	No	No	Yes
Emotions for Children	Android	No	No	Yes	Yes
My Emotions	Android	No	No	No	Yes
Learning emotions	Android/iOS/				
Windows	No	No	No	Yes	
Montessori	Android/iOS	No	No	No	Yes
Emotions Project	Android/Windows	No	No	Yes	Yes
PictoTEA	Android	No	No	No	Yes
MITA	Android/iOS	No	No	No	Yes
Autism-Discovering emotions	Android/iOS	No	No	Yes	Yes
How Do I Feel?	Android	No	No	Yes	Yes
Emotion Learning	Android	No	No	No	Yes
#Autism Emotions	Android/iOS	No	No	Yes	Yes
Toddler Feelings	Android	No	No	No	Yes
José Aprende	Android/iOS	No	No	No	Yes
Even better	Web	No	No	No	Yes

try to use the children's usually intact systemizing abilities to their benefit to help them learn how to look at faces and recognize expressions on them. Another example of a study using multimedia resources is [6], which uses the Mindreading DVD to teach children how to recognize emotions. This DVD is essentially a set of 412 emotions, with each emotion being expressed by 6 different actors and actresses of different ages and cultures.

Our proposal covers some of the gaps found in these applications since it includes a mechanism to automatically check the progress of the children with immediate feedback thanks to NFC and *Affectiva* technologies. The system proposed in this paper is a serious game which has been developed for Android devices and uses both NFC and emotion detection (*Affectiva*) technologies. Furthermore, the picture exchange communication system (PECS) has been embedded in the activities of each game implemented in our proposal. As we indicated above, the key contribution of this proposal lies in the integration of different technologies, namely automatic emotion detection and tangible user interfaces, in order to provide an easy-to-use serious game so that therapists working with children with ASD can teach them to recognize emotions with a tool that keeps them engaged and entertained. In the next section, we present a description of the software application developed.

3 System description

EmoTEA is a serious game developed as a mobile application designed to help children with ASD to improve and develop their emotional intelligence, especially emotional skills regarding emotion recognition, whether their own emotions or the ones expressed by someone else. The system has two separate and well-defined parts: firstly, a user management section aimed at the person in charge of those children with ASD (a psychotherapist, their parents or legal tutor, etc.), and secondly, a section which contains the actual games, which is the part the users access after their carer has registered them in the application.

Regarding the multimedia resources used in the development of EmoTEA and shown in the figures used throughout this paper, all the images and videos were taken from [2, 38, 41] and [50], with license of use.

EmoTEA can be defined via three main elements:

- *Target population*. The application is aimed at children with ASD aged between 6 and 12 years old, and this age range was set in agreement with the Association "Autism Development." It is also worth mentioning that the idea of EmoTEA itself arose from a collaboration with said association;
- *Technology*. EmoTEA has been developed using tangible user interfaces as the main interaction mechanism, and automatic face-based emotion detection. Besides using Android as the development platform, Affectiva's SDK and NFC tags were used to enable EmoTEA to recognize emotions on the basis of facial expressions, and to turn physical objects into interfaces the app can interact with, respectively. By including TUIs in the application we take advantage of their beneficial effects in learning settings, while Affectiva gives EmoTEA the power to automatically evaluate users in emotion mimicking games;
- *Target Skills*. As stated above, one of the difficulties people with ASD face is the inability to express and/or understand emotions, and to recognize their own ones or the ones expressed by other people. EmoTEA's main purpose is to tackle this problem, offering exercises to learn how to identify and express basic emotions [15]. These exercises, based on emotion identification and mimicking, seek to help people with ASD to develop their emotional intelligence skills.

The idea of developing this application arose from the collaboration established with the local Association called "*Autism Development*" and their need to improve the emotional skills of children with autism spectrum disorder. In the course of several meetings, the main requirements of EmoTEA were established, such as the target population, for instance. By performing a critical review of studies on autism, it was observed that the authors in [10] concluded that autism spectrum disorders were usually diagnosed at ages which range from 3 to 10 years old, approximately, with that range shrinking by three years, from 3 to 7 years old, for children with autistic disorders. Furthermore, in [29] the authors state that ASD diagnoses are usually quite stable, i.e., diagnoses made in the early years of the child are rarely mistaken. The authors in [23] analyzed the progress of several children with ASD from their early childhood to their early adolescence, concluding that even when most of the participants did not experience any improvement, there was a small group of them who did. Based on this information, the target population was chosen to be children with ASD aged between 6 and 12 years old, since it is a range of ages in which most of the diagnoses have been made but children have not entered adolescence yet, so it is easier for them to grasp new emotional skills [7],12. However, it is important to recall that ASD is a *spectrum* disorder, and the differences between each patient (intellectual ability, associated symptoms, etc.) can be huge, which may lead us to modify the limits of this range for certain situations. For instance, a child within this age range with extreme symptoms and low cognitive skills may need dedicated help from a psychologist or assistant, and would not be able to use EmoTEA. On the other hand, a child younger than 6 or older than 12 with good intellectual skills may be suitable for the app. This is why the application was developed for children with autism spectrum disorder aged between 6 and 12 years old, but this range of ages may vary.

Regarding the technical development aspects of EmoTEA, tangible user interfaces (TUIs) are included as the main interaction mechanism with the system to make it easier to use for children with special needs, based on existing evidence in the literature about how the naturalness of TUIs allows children to be more explorative and expressive [43],48. TUIs were implemented as physical cards with images representing the different emotions, with these images being both pictograms and pictures of real faces portraying some emotion. Children have to manually handle the cards to find the one required by the system. In this way, they can learn to differentiate emotions in a more playful way. In addition, affective computing technology is included in the system to support the identification of facial expressions and to help children to learn how to express emotions. In this case, they have to "imitate" the emotion depicted by a picture on the screen and in this way they learn to express it. Finally, taking into account what the children have previously learned in the first activities, they will have the chance to apply this knowledge by observing videos and identifying the emotions shown in them.

Introducing this type of applications in the teaching process of children with ASD is a challenge because it is difficult to alter their routine and when they change the activities they usually perform or the environment where they perform them, they get nervous and their behavior changes. Therefore, this application also aims to observe the adaptation of children to the changes that take place within their environment.

The technical requirements for this application to run properly are minimal. It only requires a touch-based device (mobile phone or tablet), Android O.S. version 5.1 (Lollipop) or higher, an NFC reader (available on most smartphones and tablets), an Internet connection and a camera with a minimum of 3 megapixels.

Additionally, and based on the experience we gained while assessing the system, we established the following ergonomic guidelines:

- There must be good lighting so that the users can see themselves. It is recommended to place the camera facing away from the light source and in front of the user. The user's face must be clear (with the exception of glasses as they can be worn). The users must be facing the camera and keeping their hands from their faces since the application will not detect their face if they turn their head or they are partially covering it with their hands;
- The device's camera must be placed in front of the user and focused on the face of the user, thus avoiding possible shadows that could be created if the camera is placed at a different angle. The recommended distance to place the camera at is about 30 cm from the user. If the device with the integrated camera is resting on a table, you should find a suitable chair so that the camera is placed in the correct position. The cards depicting the different emotions must be brought close to the NFC reader, to within a distance of less than 10 cm for the reader to detect them properly and allow the application to work correctly.

It is also important to point out that even though the application was developed and assessed within the context of the "Autism Development" Association, the application domain of our system is much bigger, and EmoTEA can be used in more general settings, e.g., at home, in regular classes at school, etc.

In the next subsection, we describe the three different games that make up the application, each of them with specific features and goals.

3.1 Game 1

In the first game, the user interacts with the mobile device and the cards representing the basic emotions as tangible objects. The app shows a picture of a face, and the user has to choose from among the different cards the one which represents the emotion corresponding to that picture, bringing the chosen card close to the NFC reader built into the mobile device. When they get 4 or more out of the 6 emotions right, they pass the level and, therefore, move on to the next one. If they fail, they keep playing at the same level. The application provides different types of feedback depending on whether the user succeeds or fails, always encouraging them to try again.

In the first level of difficulty, different images of real faces representing the basic emotions are shown and the user has to choose the card which corresponds to the emotion requested, as shown in Fig. 2. In this level, the cards available to the users are pictures of real faces.

Once the first level is passed, the second level is unlocked. In this second level, real images are shown on the device, as in the previous level, but in this case the cards offered to the users show pictograms representing the different emotions. In this way, the user has to choose from among the different cards which pictogram corresponds to the real image shown on the device. Therefore, they learn how to identify different representations of the same emotion, both in real pictures and in pictograms.

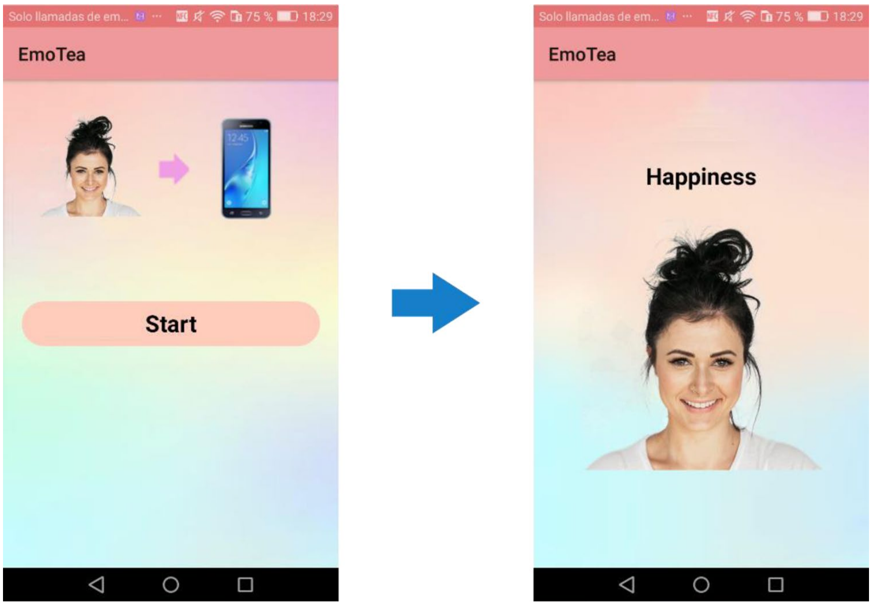

Solo llamadas de em... 75 % 18:29
EmoTea
Start
Solo llamadas de em... 75 % 18:29
EmoTea
Happiness

Fig. 2 Game 1—level 1

3.2 Game 2

Once the children have learned and identified the different emotions, linking emotions with real (faces) and conceptual (pictograms) representations of them, they are now ready to learn how to express them themselves. This is the main purpose of the second game: teaching children how to express emotions with their faces by mimicking. When the game starts, the picture of an emotion is shown, and the name of the emotion is displayed. Then, the user has to express it with their face, and their facial expressions are analyzed to know whether they are expressing the emotion correctly or not. The feedback provided in this game is the same as in the previous one. The emotion detection through facial expressions was implemented by integrating *Affectiva* technology [1] in the solution proposed.

In the first level, pictures of emotions expressed by real faces are shown. The user has to imitate the facial expression according to the expression depicted in the picture, as shown in Fig. 3.

In the second level, the goal is the same, but this time the system shows pictograms instead of real pictures to ask the user to express an emotion. It is not so much a question of "imitating" what the pictogram shows as a question of recognizing what emotion a pictogram is portraying and knowing how to express it themselves. Figure 4 shows an example of this level in which emotions of joy and anger had to be expressed.

With games 1 and 2, we aim at teaching children with ASD how to recognize and express the facial expression of the six basic emotions. As for the TUIs used in these first two games, the user requirements stated that the cards used as TUIs shall include both pictograms and real pictures, so that players could have different references to learn about feelings and how to identify them.

3.3 Game 3

In this last game, the system displays fragments of the Pixar film called "Inside out" showing situations in which the emotions learned in the previous games can be identified. In this game, users also interact with the cards, used as TUIs. The aim of this game is to help users recognize emotions in contextual situations. To this end, firstly the user watches the video fragment for a few seconds, and then they have to choose the corresponding emotion from among the cards available for this purpose, bringing the correct card close to the mobile device with the NFC reader, as in game 1.

This game has been divided into two levels of difficulty according to the difficulty of recognizing emotions. In the first level, we set joy, sadness and anger as the possible emotions to be recognized, and in the second level, we set surprise, fear and disgust, since they are more difficult to recognize than the first ones. Figure 5 shows a video fragment representing joy in a contextual situation.

This third game perfectly complements the first two games since it allows the users to put their emotion recognition skills to the test. In this game, emotion is not portrayed by a fixed image, but by a whole set of features: the faces of the characters shown in the video, their surroundings, the color palette, the theme of the scene, etc. By playing this game, users start linking emotions, not only to faces and

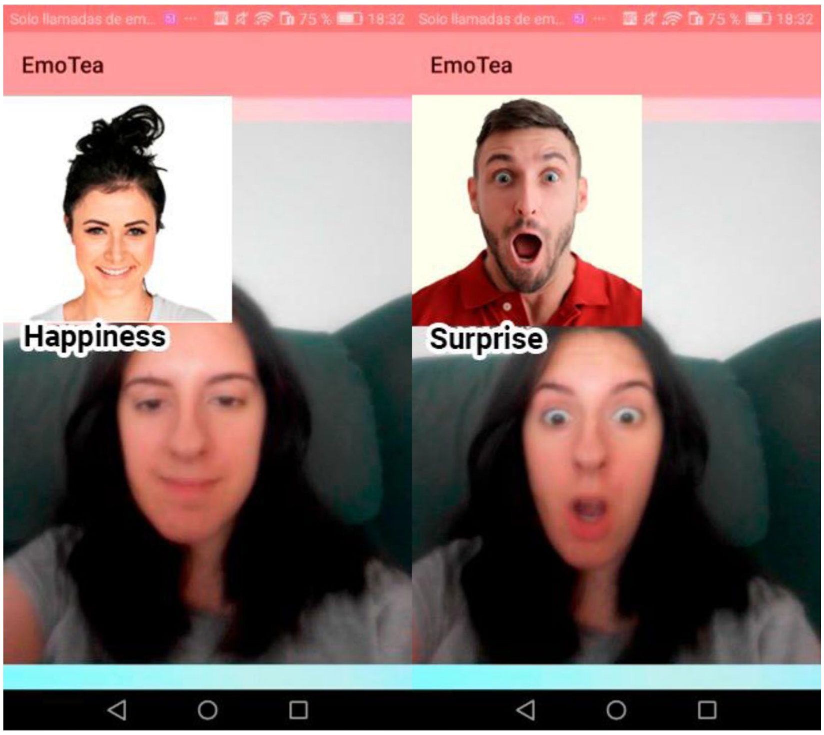

Solo llamadas de em... 75 % 18:32
EmoTea
Happiness
Solo llamadas de em... 75 % 18:32
EmoTea
Surprise

Fig. 3 Game 2—level 1. Imitating joy and surprise

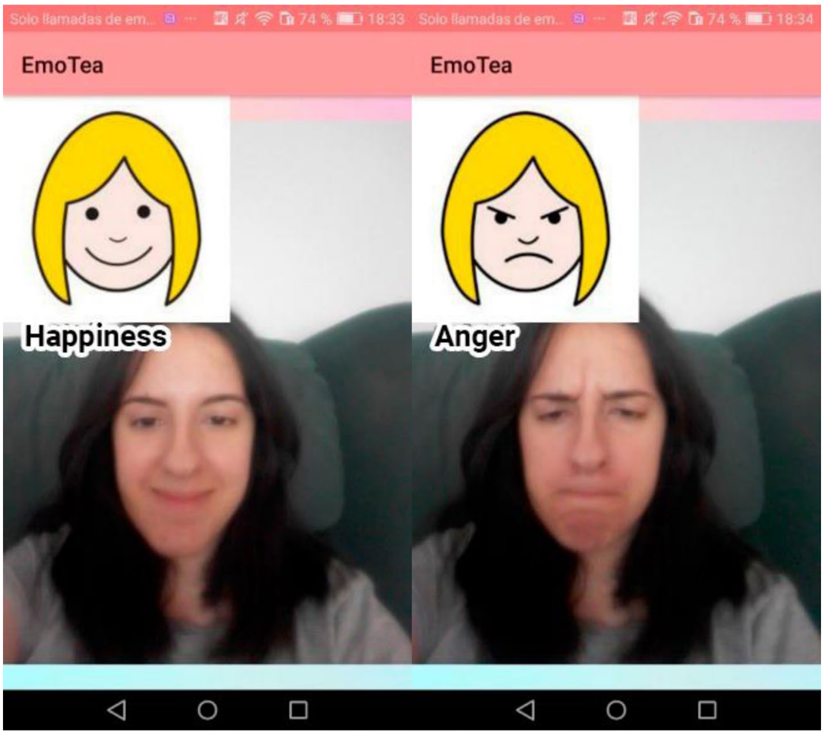

Solo llamadas de em... 74 % 18:33
EmoTea
Happiness
Solo llamadas de em... 74 % 18:34
EmoTea
Anger

Fig. 4 Game 2—level 2. Expressing joy and anger

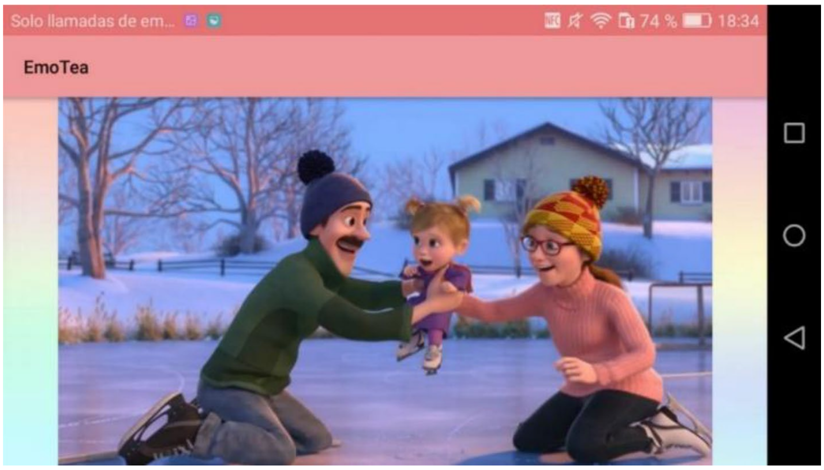


Fig. 5 Game 3—level 1

words but to situations, which helps them to generalize their emotion recognition knowledge to new settings.

4 Preliminary assessment of the system

As was mentioned above, we carried out an evaluation with ASD specialists from the Association in order to assess the teaching capabilities of the application in learning emotions, and its acceptance by the specialists who work with children with ASD. The idea was to assess the use of our system as a good alternative or complement to the traditional therapies they usually apply for this purpose, as well as its usability. The system was assessed on the premises of the Autism Development Association, which is registered by the Local Council under Organizations for Disabled People, with number 25.2255/03. This association works at helping children with ASD develop their personal skills and facilitate their integration into society. Both the children and the psychotherapists belonging to this association participated in the evaluation, and in this way the educators could assist the children with ASD while using the application in a real usage scenario.

The study was ethically approved by both the executive team and the professional staff of the Autism Development Association and was in accordance with the Declaration of Helsinki [54]. Before the beginning of the assessment sessions, informed consent was obtained from all the participants, including the parents or legal guardians of the children participating in the assessment. The children also had their parents' permission to use images or videos of the evaluation sessions.

To limit the scope of this process, the evaluation was delimited so that the participants would only play the first level of each one of the games, since the corresponding second levels use the same mechanics. Below we describe the key aspects considered during the assessment:

- *Functionality*. The application provides what is necessary to successfully meet the objectives for which it was designed;
- *Usefulness*. Educators believe that the application is useful to improve the emotional skills of children in a playful and entertaining way;
- *Ease of use*. The application has a simple interface both for those who have been previously informed about how it is used and for those who do not have prior information.

In summary, the preliminary assessment performed was mainly focused on the acceptance of the system by the specialists who work with children with ASD, as a good alternative or complement to the traditional therapies they usually apply to teach children emotion-related concepts. This preliminary assessment helped us gather their opinions on the usefulness of the system and identify specific usability problems.

4.1 Method

In this section, we present the method used during the assessment, including the description of participants, the context in which the evaluation was carried out, the tasks to be performed, as well as the place, the device used to run the application and the tools used by the evaluators. We also describe the process and the usability metrics applied. Finally, we present the results of the evaluation.

4.1.1 Evaluation techniques

The system evaluation was designed considering traditional evaluation techniques [13] and the current context of this project. The main goal of this preliminary evaluation, as has been mentioned above, was to assess the suitability of the proposed tool for its objective, that is, to teach children with ASD to recognize emotions. To this end, we designed the evaluation by combining the techniques of *Cognitive Walkthrough, Thinking Aloud* and *Cooperative Evaluation*. The evaluation was carried out by specialists from the Association and one child each time, so the psychotherapists could check in real time how the children react to the application, whether they understood how it works, whether the content was appropriate for the skill they should learn from it, whether they were getting frustrated without us noticing, etc. During the whole process, the specialists made comments to the evaluator about the children's reactions, the application's usability, etc. One of the benefits of this evaluation technique is that we do not need many children to assess the tool, but just a set of archetypical participants who are representative of the different kinds of children that might use the system.

4.1.2 Participants

The evaluation was carried out with two psychotherapists and three children of different ages and different degrees of autism. The children, aged between 8 and 10 years, had some previous experience of using similar applications and devices such as tablets and/or mobile phones. One of these children presented a mild degree of autism, while the other two presented a more severe degree. This sample allowed us to appreciate differences between users with different levels of ASD. Despite the apparent small number of children we could recruit, as the main goal was to assess the acceptance of the system by the physiotherapists, we designed the evaluation as a cognitive walkthrough, where it is more important to have representative participants than to have a lot of children with similar characteristics repeating the evaluation tasks. Finally, at the end of the evaluation, the psychotherapists who participated in the evaluation provided important feedback, as experts in the field, in relevant aspects observed during the evaluation. Both the educators and the children, with their help, completed the SUS (System Usability Scale) questionnaire to measure the users' satisfaction [9].

4.1.3 Context of use

We defined two tasks to be performed during the evaluation: navigate between the different parts of the application and play the first level of each game. Although they are presented here as two different tasks, the navigation task was transversal to the playing task. The navigation task consists in guiding the user through the different screens and levels of the application. It is important to note that the evaluation was not focused on evaluating the children's knowledge of emotions, but their ability to interact with the application, that is, whether they were able to interact properly with the cards as TUIs and express emotions in front of the mobile camera in a natural way. The tasks that the participants had to perform were the same as those that a user willing to complete the entire game should carry out. The navigation task was considered finished when the "Thank you for playing" screen popped up, while the task of playing the first level of each game was considered finished when the "Level passed" screen popped up.

As mentioned at the beginning of the section, the evaluation was carried out on the premises of the Autism Development Association of Albacete. The evaluation was performed individually, i.e., there was one child with the psychotherapists in each evaluation session. The evaluator was in charge of writing down the most important aspects of the evaluation, promoting the Thinking Aloud technique to gather all comments and suggestions from the educators, as well as helping children in the case of technical problems. We used a stopwatch to calculate the time spent performing each task, and a camera to take pictures and videos during the evaluation (see Fig. 6). Finally, all the participants completed the SUS questionnaire at the end of the evaluation.

4.1.4 Experimental design

Before starting the evaluation, the parents of the participants were required to sign an authorization to allow their children to participate in it. When children arrived at the assessment session, they were informed about the evaluation process for testing the application. They were also informed that this evaluation would only be used for testing the software

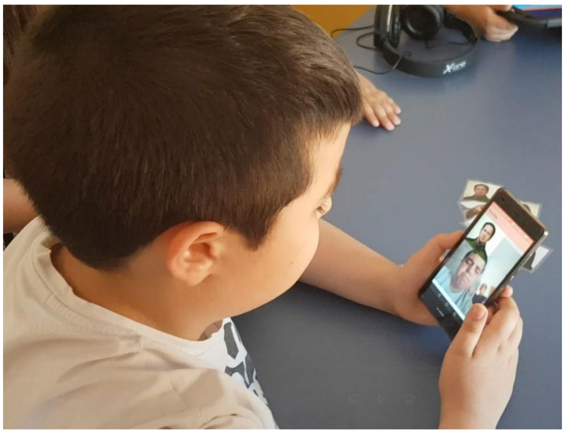


Fig. 6 Child interacting with EmoTEA during the evaluation process

application, not for assessing their personal capabilities. Initially, children were informed that the evaluation consisted in playing the first level of the three games, and involved interacting with the cards (tangible objects) and the device's camera. The interaction mechanisms with both the cards and the camera were also explained in detail. Once all the above were completed, we began with the individual evaluation of each participant together with their psychotherapist.

At the end of the evaluation, all participants were asked to complete an adapted SUS questionnaire. Before completing it, they were instructed about how to do so. The educators also completed the SUS questionnaire to give feedback from their point of view as specialists in the field.

The tasks performed in the evaluation were the following:

- *Task 1: Browse the application*. This task was performed throughout the evaluation and consisted in browsing the application to test that the navigability and options provided were easily understood. In this task, the user was guided through the different screens of the application, and this represented the flow a user would follow while using the application, except for the fact that the participants of this evaluation only played one level of each game;
- *Task 2: Performing the first level of the three games*. To simplify the evaluation, children only had to tackle the first difficulty level of each game. This is enough for them to interact with the tangible objects and the device's camera to test the practicality and ease of use of these interaction mechanisms incorporated in the system to play the different games.

4.1.5 Usability metrics

The usability metrics applied in the evaluation were the following:

- Effectiveness
    - Completion rate. Percentage of tasks completed with (assisted completion rate) and without (unassisted completion rate) the help of the person in charge of the evaluation;
    - Errors. Actions that do not lead to completing the task or times the child needs to tackle the task in order to complete it;
    - Assistance. The number of times that help is offered by the person in charge of the evaluation so that a task can be carried out and finished;
- Efficiency
    - Task time. Amount of time, in minutes and seconds, needed for a user to complete a task;
- Satisfaction. This metric was measured with the System Usability Scale Test (SUS test) that all participants completed at the end of the assessment session. The questionnaire consists of 10 questions that assess various aspects related to the usability of the application [9].

4.2 Assessment results

Table 2 shows the outcomes of the first task involving browsing the application, whereas Table 3 shows a statistical summary of these data. In addition, we can observe that, except for one participant, the others were able to complete the tasks on their own.

The outcomes of the second task show that the participants required more help in this case. Table 4 shows the outcomes of this task, and Table 5 shows the statistical summary of these data.

Finally, the average value obtained from the SUS test was 90.625 out of 100. Apart from the answers to the SUS test, it was important to know the opinion and suggestions of the educators in order to obtain information that might be valuable for the further improvement of the system, and assess the acceptance of the system, so we wrote down all the comments they made on different aspects that arose during the assessment.

5 Assessment discussion and conclusions

The evaluation process was carried out without major problems, and the results obtained provided very valuable information. The browsing activities presented no problems for the participants, as the educators informed us that the children had previous experience with computer devices. However, the activities involving playing the different games posed a bigger challenge for our participants. For instance, the children with a more severe degree of autism had some problems when expressing emotions using their facial expression (game 2). These problems are conveyed by the increasing assisted completion rate, errors, and assistance

Table 2 Results of task 1 (browsing the app)

Participant	Assisted completion rate (%)	Unassisted completion rate (%)	Time	Errors	Assistance
1	0	100	3:45	0	0
2	0	100	4:43	1	0
3	14.29	85.71	5:02	1	1

Table 3 Statistical data for Task 1

Statistical Values	Assisted completion rate (%)	Unassisted completion rate (%)	Time	Errors	Assistance
Mean	4.76	95.24	4:30	0.67	0.33
Min	0	85.71	3:45	0	0
Max	14.29	100	5:02	1	1
Standard deviation	8.25	8.25	0:40	0.58	0.58

Table 4 Results of Task 2 (playing games)

Participant	Assisted completion rate (%)	Unassisted completion rate (%)	Time	Errors	Assistance
1	85.71	14.29	3:00	0	1
2	71.43	28.57	4:09	1	2
3	71.43	28.57	4:32	0	2

during the playing task (Table 5). The use of NFC tags also posed a challenge at the beginning of the evaluation: one of the children tried to place the cards over the screen, since he did not know where the NFC reader was. Surprisingly, they very quickly learned the interaction mechanism of grasping the cards and bringing them close to the mobile device. Despite these initial difficulties, the children became totally engaged with the application and enjoyed its interaction mechanisms (TUIs), since they were new and fun for them. As a matter of fact, they wanted to continue playing after having finished the session. The psychotherapists highlighted this fact, as children usually get tired very soon when doing any kind of training, but in this case, their attitude was totally different. Thus, one of the main findings of the assessment was the engagement of the children when using the system.

As for the SUS test results, they fit with what the educators told us about the application with the Think Aloud assessment technique. Once the children learned how to use the cards (as TUIs), they started to have fun, even when some of the participants could not finish the second game. We will consider all these data when preparing a future version of the application.

In addition, it is important to acknowledge the limitations of the evaluation performed. It was carried out in a very controlled environment, with each child being assisted by their psychotherapist. A more complete, long-term evaluation with a higher number of participants is still necessary to really assess the educative value of the system proposed and to detect further usability issues.

In conclusion, the psychotherapists who participated in the system evaluation gave their approval and considered the system a good tool to teach emotion-related concepts to children with ASD. One of the main features that contributed to the success of the system and that differs from previous works is the combination of automatic emotion recognition and tangible objects as the main interaction mechanism. According to them, the system not only has the necessary contents to teach the different aspects of emotion recognition to children, but also has an innovative interaction mechanism, based on tangible and graspable cards, which appeals to children and keeps them engaged for longer while they are also learning. This is one of the most common deficiencies in applications of this nature according to the existent literature, since researchers and developers usually overlook the interaction mechanisms integrated in their systems in favor of creating more different types of exercises, more complex control panels, etc.

Some improvements are proposed as future work based on the collaboration established with experts in this field. Firstly, the order in which the pictures currently appear is always the same. It seems that a random order would be a much better idea to prevent children from learning the order in which the pictures appear, and thus learning the emotion by heart, not by a correct identification. On the basis of a similar concept, we also plan to increase the number of real pictures, as looking at a wider range of different people will help them identify the different expressions of an emotion, since not all people express emotions in the same way. Secondly, we also plan to include

Table 5 Statistical data for Task 2

Statistical values	Assisted completion rate (%)	Unassisted completion rate (%)	Time	Errors	Assistance
Media	76.19	23.81	4:20	0.33	1.67
Min	71.43	14.29	3:00	0	1
Max	85.71	28.57	4:32	1	2
Standard deviation	8.25	8.25	0:48	0.58	0.58

collaborative activities. Collaboration in reaching different levels would encourage them to practice other important abilities. Lastly, secondary emotions could be added so that they could learn to identify more sophisticated and complex types of emotions.

6 Conclusions

Autism spectrum disorder is a neurodevelopmental disorder which impairs the social skills of a person, especially those relating to emotional awareness and emotion recognition. However, these emotion-related skills can be learnt, especially if this learning process starts in early childhood.

In this paper, we present a novel system based on tangible user interfaces implemented with NFC technology and face-based emotion recognition software to help children suffering from ASD recognize and express emotions, supporting our proposal on the existent literature related to specialized therapies for children with this disorder. Our application is mainly based on novel interactive mechanisms, namely automatic emotion recognition through the device's built-in camera, and tangible user interfaces (TUIs). NFC (near field communication) technology has also been used to implement natural interfaces for children to handle the objects needed for playing the different games. TUIs provide a familiar and simple way for children to interact with the game in a fun and intuitive way. According to the specialists, we have developed a serious game that avoids disruptive elements so that the attention of children can be focused on learning how to identify emotions in different situations as well as how to express such emotions themselves. The software application has been assessed with children with ASD and their psychotherapists in a real setting, obtaining very good results and the specialists' acceptance of the system as a useful tool to be used for teaching emotion-related concepts. In addition, we have gathered important feedback that will help us improve the application. Once these improvements have been implemented, we plan to carry out a long-term evaluation of our tool to assess its impact on learning.

Acknowledgements This work has been partially supported by the national project granted by the Ministry of Science, Innovation and Universities (Spain) with reference RTI2018-099942-B-I00, and by the project TecnoCRA (ref: SBPLY/17/180501/000495) granted by the Regional Government (JCCM) and the European Regional Development Funds (FEDER). We would also like to thank Mar Cabañero, undergraduate student, for her participation in the implementation and testing of the software application.

Declarations

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

References

1. Affectiva (2018) Affectiva. Emotion Recognition Software and Analysis. Retrieved from: https://www.affectiva.com/
2. ARASAAC (2018) Portal Aragonés de la Comunicación Aumentativa y Alternativa. Retrieved from: http://www.arasaac.org/
3. Artoni, S., et al.: Technology-enhanced ABA intervention in children with autism: a pilot study. Univers. Access Inf. Soc. **17**(1), 191–210 (2018). https://doi.org/10.1007/s10209-017-0536-x
4. American Psychiatric Association. Autism Spectrum Disorder In Diagnostic and statistical manual of mental disorders (5th ed.). (2013). https://doi.org/10.1176/appi.books.9781585629992
5. Baron-Cohen, S., Golan, O., Ashwin, E.: Can emotion recognition be taught to children with autism spectrum conditions? Philo. Trans. R. Soc. B: Biol. Sci. **364**(1535), 3567–3574 (2009). https://doi.org/10.1098/rstb.2009.0191
6. Baron-Cohen, S., Golan, O., Wheelwright, S. & Hill, J.J.:Mind Reading: the interactive guide to emotions. London, UK: Jessica Kingsley Limited, (2004) (http://www.jkp.com/ mindreading)
7. Bölte, S.: "Is autism curable?," Developmental Medicine and Child Neurology, vol. 56, no. 10. Blackwell Publishing Ltd, pp. 927–931, 01-Oct-2014, https://doi.org/10.1111/dmcn.12495.
8. Bondy, A.S., Frost, L.A.: The picture exchange communication system. Focus Autistic Behavior **9**(3), 1–19 (1994)
9. Brooke, J.: SUS-A quick and dirty usability scale. In: Jordan, PW., B. Thomas, B Weerdmeester, Bernard A and McClelland, IL (eds) Usability evaluation in industry, 1st edn. Great Britain Taylor and Francis, pp. 189–194, (1996)
10. Daniels, A.M., Mandell, D.S.: Explaining differences in age at autism spectrum disorder diagnosis: a critical review. Autism **18**(5), 583–597 (2014). https://doi.org/10.1177/1362361313480277
11. Daou, N., Hady, R. T., & Poulson, C. L.: Teaching children with autism spectrum disorder to recognize and express emotion: A review of the literature. Int Elect J Elemen Educ **9**(2), 419–432 (2016).
12. G. Dawson, K. Zanolli, Skuse, C. Frith, and Schultz, "Early intervention and brain plasticity in autism," Novartis Found. Symp., vol. 251, pp. 266–280, 2003, https://doi.org/10.1002/0470869380.ch16.
13. Dix, A., Finlay, J., Abowd, D., G. and Beale R.: Human-Computer Interaction, 3rd edn., pp. 318–364. Prentice-Hall Europe, Harlow, England (2004)
14. Djaouti, D., Alvarez, J., Jessel, J. P.: Classifying serious games: the G/P/S model. In: Handbook ofresearch on improving learning and motivation through educational games: Multidisciplinary approaches (pp. 118–136). IGI Global (2011)
15. Ekman, P.: Basic emotions Handbook of cognition and emotion, pp. 45–60. Wiley (1999)

16. Ekman, P., Friesen, W.V.: Measuring facial movement. Environ. Psychol. Nonverbal Behav **1**(1), 56–75 (1976)
17. Ekman, P., Oster, H.: Facial expressions of emotion. Ann Rev Psychol **30**(1), 527–554 (1979)
18. Finkelstein, S.L., Nickel, A., Harrison, L., Suma, E. A.., and Barnes T.: "cMotion: A new game design to teach emotion recognition and programming logic to children using virtual humans. In: Proceedings—IEEE Virtual Reality, 2009, pp. 249–250, https://doi.org/10.1109/VR.2009.4811039.
19. Garcia-Garcia, J.M., Penichet, V.M.R., Lozano, M.D.: Emotion detection: a technology review. In: Proceedings of the XVIII International Conference on Human Computer Interaction—Interacción**17**, pp. 1–8, (2017) https://doi.org/10.1145/3123818.3123852
20. Garcia-Garcia, J.M., Penichet, V.M.R., Lozano, M.D., Garrido, J.E., Lai-Chong Law, E.: Multimodal affective computing to enhance the user experience of educational software applications. Inf. Syst Mob (2018). https://doi.org/10.1155/2018/8751426
21. Golan, O., et al.: Enhancing emotion recognition in children with autism spectrum conditions: an intervention using animated vehicles with real emotional faces. J. Autism Dev. Disord. **40**(3), 269–279 (2010). https://doi.org/10.1007/s10803-009-0862-9
22. Google Scholar. Retrieved from: https://scholar.google.es/schhp?hl=es&as_sdt=0,5
23. Gotham, K., Pickles, A., Lord, C.: Trajectories of autism severity in children using standardized ADOS scores. Pediatrics **130**(5), e1278–e1284 (2012). https://doi.org/10.1542/peds.2011-3668
24. Granic, I., Lobel, A., Engels, R.C.: The benefits of playing video games. Am. Psychol. **69**(1), 66 (2014)
25. Hansen, O.B., Abdurihim, A., and McCallum, S.: Emotion recognition for mobile devices with a potential use in serious games for autism spectrum disorder. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2013, vol. 8101 LNCS, pp. 1–14, https://doi.org/10.1007/978-3-642-40790-1_1.
26. Jack, R.E., Garrod, O.G.B., Schyns, P.G.: Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. Curr. Biol. **24**(2), 187–192 (2014). https://doi.org/10.1016/j.cub.2013.11.064
27. Jain, S., Tamersoy, B., Zhang, Y., Aggarwal, J. K., and Orvalho, V.: An interactive game for teaching facial expressions to children with autism spectrum disorders. In: 5th International Symposium on Communications Control and Signal Processing, ISCCSP 2012, (2012), https://doi.org/10.1109/ISCCSP.2012.6217849.
28. Kleinginna, P.R.J., Kleinginna, A.M.: A categorized list of emotion definitions, with suggestions for a consensual definition. Motiv. Emot. **5**(3), 263–291 (1981)
29. Lord, C., Risi, S., DiLavore, P.S., Shulman, C., Thurm, A., Pickles, A.: Autism From 2 to 9 Years of Age. Arch. Gen. Psychiatry **63**(6), 694 (2006). https://doi.org/10.1001/archpsyc.63.6.694
30. Miranda, J. C., Fernandes, T., Sousa, A. A., Orvalho, V.: Interactive technology: teaching people withautism to recognize facial emotions. Autism Spectrum Disorders-From Genes to Environment. Prof. Tim Williams(Ed.) InTech ISBN: 978-953-307-558-7, (2011). pp. 299–312
31. Messinger, D. S., Duvivier, L. L., Warren, Z. E., Mahoor, M., Baker, J., Warlaumont, A., Ruvolo, P.: affective computing, emotional development, and autism. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of affective computing (pp. 516–536). Oxford University Press (2015). https://doi.org/10.1093/oxfordhb/9780199942237.013.012.
32. Miranda, J.C., Fernandes, T., Augusto, A., and Orvalho, V.C..: Interactive Technology: Teaching People with Autism to Recognize Facial Emotions. In: Autism Spectrum Disorders—From Genes to Environment, InTech, 2011.
33. Noor, H., Shahbodin, F., Pee, N.: Serious game for autism children: review of Literature. WorldAcademy of Science, Engineering and Technology, Open Science Index 64. Int J Psychol Behav Sci **6**(4), 554–559 (2012)
34. Mohzan, M.A.M., Hassan, N., Halil, N.A.: The influence of emotional intelligence on academic achievement. Proc. Soc. Behav. Sci. **90**, 303–312 (2013). https://doi.org/10.1016/j.sbspro.2013.07.095
35. Mondragon, A.L., Dufresne A., Nkambou, R., and Poirier P.: An Affective Intelligent Tutoring System In The Special Education Of Individuals With Autism. In: EDULEARN17 Proceedings, 2017,**1**, pp. 4114–4122, https://doi.org/10.21125/edulearn.2017.1884.
36. Nisiforou, E.A., Zaphiris, P.: Let me play: unfolding the research landscape on ICT as a play-based tool for children with disabilities," Universal Access in the Information Society,**19**, no. 1. Springer, pp. 157–167, 01-Mar-2020, https://doi.org/10.1007/s10209-018-0627-3.
37. Núñez Castellar, E., Van Looy, J., Szmalec, A., De Marez, L.: Improving arithmetic skills through gameplay: assessment of the effectiveness of an educational game in terms of cognitive and affective learning outcomes. Inf. Sci. (Ny) **264**, 19–31 (2014). https://doi.org/10.1016/j.ins.2013.09.030
38. Pexels (2020) Pexels: Fotos de stock gratis. Retrieved from: https://www.pexels.com/es-es/
39. Piana, S., Staglianò, A., Camurri, A., Odone, F.: A set of full-body movement features for emotionrecognition to help children affected by autism spectrum condition. In: 1st International Workshop on Intelligent Digital Games for Empowerment and Inclusion (IDGEI) (2013)
40. Picard, R.W.: Affective computing. MIT Press **321**, 1–16 (1995). https://doi.org/10.1007/BF01238028
41. Pixar.: (2018) Pixar Animation Studios. Retrieved from: https://www.pixar.com/#pixar home
42. Plutchik, R.: The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American Scientist. Sigma Xi The Scientific Research Honor Society. (2001). https://doi.org/10.2307/27857503
43. Rante, H., Lund, M., Caliz, D.: (2018) The Role of Tangible Interfaces in Enhancing Children's Engagement in Learning. In: International Conference on Electrical Systems, Technology and Information (ICESTI 2017).
44. Reichow, B., Steiner, A.M., Volkmar, F.: Social skills groups for people aged 6 to 21 with autism spectrum disorders (ASD). Campbell Syst. Rev. **8**(1), 1–76 (2012). https://doi.org/10.4073/csr.2012.16
45. Salovey, P., Mayer, J.D.: Emotional Intelligence. Imagin. Cogn. Pers. **9**(3), 185–211 (1990). https://doi.org/10.2190/DUGG-P24E-52WK-6CDG
46. Schuller, B., Marchi, E., Baron-Cohen, S., O'Reilly, H., Pigat, D., Robinson, P., Daves, I.P.: The state of play of ASC-Inclusion: an integrated Internet-based environment for social inclusion of children with autism spectrum conditions (2014). arxiv:1403.5912
47. Sim, L., Zeman, J.: Emotion awareness and identification skills in adolescent girls with bulimia nervosa. J. Clin. Child Adolesc. Psychol. **33**(4), 760–771 (2004). https://doi.org/10.1207/s15374424jccp3304_11
48. Shaer, O., Hornecker, E.: Tangible user interfaces: past, present and future directions. Found. Trends® Human-Comput. Interact. **3**(1–2), 1–137 (2010)

49. Tsai, T.W., and Lin, M.Y.: An application of interactive game for facial expression of the autisms. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 6872 LNCS, pp. 204–211, (2011) https://doi.org/10.1007/978-3-642-23456-9_37.
50. Unsplash (2020). Unsplash: Beautiful Free Images & Pictures. Retrieved from: https://unsplash.com/
51. WebMD. What Are the Treatments for Autism? Retrieved from: https://www.webmd.com/brain/autism/understanding-autism-treatment
52. Wierzbicka, A.: Human emotions: universal or culture-specific? Am. Anthropol. **88**(3), 584–594 (1986)
53. Wikipedia. (March 27th, 2019). "Robert Plutchik" Retrieved from: https://es.wikipedia.org/wiki/Robert_Plutchik
54. World Medical Association. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human sub-jects. Bulletin of the World Health Organization **79** (4), 373–374 (2001). World Health Organization. https://apps.who.int/iris/handle/10665/268312
55. Yan, Y., Liu, C., Ye, L., Liu, Y.: Using animated vehicles with real emotional faces to improve emotion recognition in Chinese children with autism spectrum disorder. PLoS ONE (2018). https://doi.org/10.1371/journal.pone.0200375
56. Yeni, S., Cagiltay, K., Karasu, N.: Usability investigation of an educational mobile application for individuals with intellectual disabilities. Univers Access Inf. Soc (2019). https://doi.org/10.1007/s10209-019-00655-0
57. Zakari, H.M., Ma, M., Simmons, D.: 2014,A review of serious games for children with autism spectrum disorders (ASD). In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8778, pp. 93–106 https://doi.org/10.1007/978-3-319-11623-5_9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.3 Building a three-level multimodal emotion recognition framework

Citación: Garcia-Garcia, J. M., Lozano, M. D., Penichet, V. M. R., and Law, E. L.C. (2022). Building a three-level multimodal emotion recognition framework. Multimedia Tools and Applications, ISSN: 1573-7721. Páginas 31. https://doi.org/10.1007/s11042-022-13254-8

Índice de impacto: 2.577, Q2 (JCR 2021)

Esta publicación avala esta tesis por compendio de publicaciones.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-13254-8

Building a three-level multimodal emotion recognition framework

Jose Maria Garcia-Garcia[1] · Maria Dolores Lozano[2] · Victor M. R. Penichet[2] · Effie Lai-Chong Law[3]

Received: 11 December 2020 / Revised: 22 February 2022 / Accepted: 15 May 2022

Abstract
Multimodal emotion detection has been one of the main lines of research in the field of Affective Computing (AC) in recent years. Multimodal detectors aggregate information coming from different channels or modalities to determine what emotion users are expressing with a higher degree of accuracy. However, despite the benefits offered by this kind of detectors, their presence in real implementations is still scarce for various reasons. In this paper, we propose a technology-agnostic framework, HERA, to facilitate the creation of multimodal emotion detectors, offering a tool characterized by its modularity and the interface-based programming approach adopted in its development. HERA (Heterogeneous Emotional Results Aggregator) offers an architecture to integrate different emotion detection services and aggregate their heterogeneous results to produce a final result using a common format. This proposal constitutes a step forward in the development of multimodal detectors, providing an architecture to manage different detectors and fuse the results produced by them in a sensible way. We assessed the validity of the proposal by testing the system with several developers with no previous knowledge about affective technology and emotion detection. The assessment was performed applying the Computer System Usability Questionnaire and the Twelve Cognitive Dimensions Questionnaire, used by The Visual Studio Usability group at Microsoft, obtaining positive results and important feedback for future versions of the system.

Keywords Affective computing · Multimodal detector · Emotion recognition · Multimedia aggregator

✉ Jose Maria Garcia-Garcia
Josemaria.garcia@uclm.es

Extended author information available on the last page of the article

Published online: 06 June 2022

1 Introduction

With the first appearance of Affective Computing (AC) 25 years ago [55], and hence of automatic emotion recognition, the following set of questions arose: What channel broadcasts emotional information with the highest clarity? How are these different channels related? Are there any other mechanisms to infer someone's emotional state? Hundreds of experiments were carried out, most of them focused on finding emotion traits in facial expressions and in voices. Furthermore, some research was conducted into whether it was possible to discover how a person was feeling according to physiological data obtained from them. A later trend even studied (and still does) how to detect emotions or affective states based on behavior. For instance, Oehl et al. [51] analyzed the pressure applied on the steering wheel by drivers to detect stress and thus possible dangerous situations; Yamachi [69] studied how the movement xof the mouse could express anxiety in the subject; and Jaques et al. [31, 44] investigated how eye movement could be used to detect emotions which are important for a learning process, such as boredom or curiosity. Nevertheless, from the beginning there was another trend which looked at how emotion recognition could be greatly improved by *fusing* the results from several emotion recognizers. This type of detector is called a *multimodal emotion detector*, since it uses several kinds of emotional information to determine what emotion a person is expressing [53].

According to several emotion theories (emotions as expressions, emotions as embodiments), emotional episodes activate multiple physiological and behavioural response systems [2]. In other words, emotional expressions are revealed through several channels at the same time (face, voice, body posture, etc.), so it makes perfect sense to read these signals from those different channels simultaneously to perform a better emotion recognition. In fact, analysing input from a single modality can lead us to wrong results. For instance, facial expression and voice can be controlled by the person we are reading data from. Physiological signals cannot be consciously modified, but the information we get from them is very primal: is the person scared? Is the person excited? When a person writes something, they may not be feeling as their writing is reflecting. All these modalities provide us with information about how someone is feeling but attending to just one of them is misleading. Someone may be smiling, but their physiological signals might be low; someone might have a straight face while saying they are comfortable, but their voice can reveal they are very nervous; someone can be completely serious, and being on the edge of a panic attack; someone might be writing about being super excited about something, while being completely emotionless. All these examples have a common point: if we only pay attention to one modality of emotion expression, we might be losing information. That is the reason why multimodal detectors can be so powerful [12].

However, the mere use of a single detector of emotions in just one channel presents several challenges in itself, so multimodal (MM) systems have largely been ignored in practice. Although we can find different proposals for multimodal processing methods, multimodal frameworks, etc., there is still a gap regarding the *practical* use of these proposals. Besides, the variety of services in existence [19] for the detection of emotions from one channel or another makes it difficult for people to easily apply them, not to mention to integrate them all together.

Not only is choosing and *acquiring* or implementing these emotion recognition services a complicated task, but also making them work together presents intrinsic challenges. Having several emotion detection services working at the same time means that we have to spend additional time making them work *together*, as we have to wait until we have received data from all of them, analyse their behaviour and the data they produce, fuse these data correctly so the aggregation of the information produces more information, instead of hiding or degrading that which we already have, etc. This new challenge that arises when we try to use different emotion recognition services together is one of the reasons why the presence of multimodal detectors is so scarce in real-world applications (or scarcer than what might be expected given their recognized superiority over unimodal emotion detection systems) [49].

Another difficulty that researchers have to face when they wish to develop a multimodal system is the *lack of references*: the proposals available in the literature or in repositories such as Github are either very abstract (proposals of UML diagrams of an ideal multimodal system) or very specific (rigid systems prepared for a certain problem or necessity). This fact defines a gap in the available resources for developers and researchers who may have operational emotion detectors but who do not have the resources to develop a system which aggregates them altogether from scratch.

In the present work, we proposed a framework called **HERA (Heterogeneous Emotional Results Aggregator)** in order to try to fill the gaps in the existent literature regarding the lack of actual implementations of frameworks. In Fig. 1 we can see a simplified diagram showing the current situation regarding the integration of heterogeneous emotion detection technologies. This figure highlights the incompatibility problems derived from this integration and how the HERA framework would fit in this ecosystem to solve these issues.

HERA offers an architecture to manage different emotion detectors producing results in different formats and to aggregate these heterogeneous results into a single, richer result. As a proof-of-concept, we have modelled and implemented our proposed framework using JavaScript and the ExpressJS framework, in the form of a web server, so it is easier to integrate multimodal emotion detection in an existing project regardless of the technology that the researchers are already using. Concerning the data fusion, HERA also contains a proposal for a *three-level data fusion model* in which data is aggregated in three different phases. Firstly, each

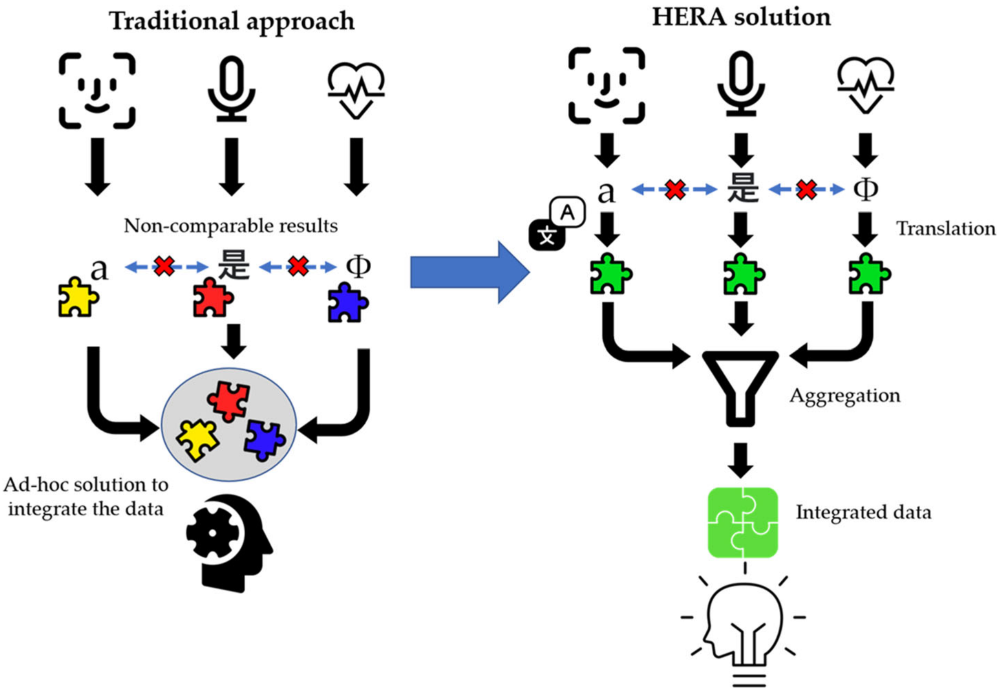


Fig. 1 Conceptual idea of the framework HERA

individual detector fuses all the data already gathered from previous requests. Secondly, detectors of the same type fuse their data together, using a data fusion strategy appropriated to these types of data. Once each group of detectors has aggregated its data and produced a single result, these data are aggregated again using a second strategy, one that fits this new context (fusing data coming from different types of affective channels and which can present contradictory values, for instance). Thus, we fuse the data in a sensible way, correcting possible discrepancies by fusing data from detectors of the same type and then being able to find new information when we fuse aggregated data considering its origin and circumstances. This aggregation process also involves a translation process, through which the heterogeneous results are translated into a common notation. For this purpose, we have used the PAD model [49] to represent the results obtained by each detector, so we store every result in the same format, regardless of the original format it had.

As it was mentioned above, this first implementation of our framework has been developed using modules and an interface-based programming approach, so it is easy to extend and adapt to particular cases. Thanks to this development approach, extending HERA is a matter of implementing new modules, without having to modify its core module at all. Also, thanks to its server nature, HERA can be deployed on any device capable of running Node (or Docker), so it can be integrated in an existent project with a minimal impact on the project's codebase.

Since the framework we propose is not only conceptual, but has been actually implemented in JavaScript as a tool for other developers, from now on we will refer to HERA indistinctly, with any of these terms: "framework", "tool", server or "system", depending on the aspect being highlighted.

It is important to point out that the key contribution of our work relies in the JavaScript implementation of the framework proposed (along with the modular structure and the data fusion framework proposed within HERA) and in the fact of offering a tangible tool to developers working in AC. While the problem of multimodality and emotion detection models has been tackled many times from many perspectives in literature before, we failed to find a proposal which actually offered an implementation for developers to easily integrate into their projects. Despite finding several frameworks which could fill the aforementioned gap in literature, they still present some deficiencies that we have tried to fix with our proposal. Since this is the goal of this proposal, the development of new data fusion algorithms or new forms of automatic classification is out of the scope of this paper.

In this paper, which is divided into six sections, we present a novel tool for integrating several emotion recognition services in the same place using a modular, scalable, fully customizable, and three-level multimodal system. In Section 2 we look at key aspects of multimodal systems, such as the kind of fusion they can use or the different systems that can be used to represent emotions. In Section 3, we present the implemented system and its different parts. In Section 4, we review the evaluation process of our multimodal system, together with the data we harvested during this process. Finally, we present several conclusions and future work in Section 5.

2 Background concepts and related works

The concept of multimodality already existed in HCI before Affective Computing came into being. Even when the technology to create these multimodal interactive systems was not yet available, HCI researchers were already testing how users preferred to interact with a system, simulating multimodality using Wizard-of-Oz systems, i.e., systems which seem to work on their own but, in reality, have 'a researcher (the "wizard") who simulates the system responses from behind the scenes' [45]. Early studies such as [28] discovered, back in 1991, that users preferred to interact through several channels at the same time. In [28], a study using a Wizard-of-Oz prototype was carried out in order to discover how the users preferred to perform a manipulation task in a three-dimensional space. Almost 60% of the participants preferred to interact with the system using both gestures and speech. This same phenomenon was described some years later in [50], where 95% of the users involved stated a preference for multimodal interaction.

When Affective Computing was defined for the first time in 1995 [54], the existent signal and data fusing knowledge became the perfect breeding ground for multimodal affective systems, i.e., systems that fuse data coming from different emotional channels (facial expression, voice tone, body gestures, electrodermal activity, etc.) to obtain a more accurate measurement, to correct anomalous samples from defective sensors, to detect affective states or behaviours that cannot be detected by attending to just one emotional channel, etc. In 1998, Chen et al. [8] implemented De Silva's proposed algorithm [61] to classify emotions expressed in 36 clips of synchronized audio/video, hence building the *first actual multimodal human emotion recognition system*. This marked the beginning of a new line of work in Affective Computing: *Multimodal Emotion Recognition (MMER)*.

This line of research has enormously advanced over the last years, taking advantage of the power that machine learning and big data offers to AC, what has led to the appearance of very different proposals to produce multimodal detectors using machine learning [11, 47, 63, 70–72]. While these works provide interesting approaches, the key contribution of this paper is to offer a software tool addressed to developers so that they can easily integrate these trained detectors into a bigger system, and this is the reason why the framework HERA does not include any type of machine learning algorithms in its implementation. Nevertheless, HERA's architecture has been designed to allow developers to ingrate their own automatic classifiers and models as local emotion detection services which behave just as the remote ones. In addition, although HERA does not use any form of machine learning to integrate data from different sources at this time, this could be included in future versions.

In the next subsection, we review how the data fusion is carried out, the different levels at which it can be performed and how these data can be represented. Lastly, we review the situation of multimodal affective technology today.

2.1 Types of data fusion

One of the inherent challenges in Multimodal Systems is *data fusion*. While being one of the key assets of MMER systems, data fusion is also one of the most difficult aspects to implement. In order to obtain added value from aggregating data from different sources, it is mandatory to combine them properly. In other words, you cannot just, for instance, calculate the average of all the data your emotion recognizers produce, but you must fuse these data following a fusion strategy. In the field of data fusion, there are *several fusion strategies*, each one taking place at a different point of the data processing flow [4, 43, 55]:

- **Raw-level fusion**. A system performs a raw-level data fusion when it works with the data at their purest level, i.e., as they are produced by the different sensors. Arrays of positions of each Facial Unit, numbers expressing the skin conductivity or temperature, the heart rate, an audio file: these data are raw data. One drawback of this kind of fusion is that sensors must be measuring the same magnitude in order to be able to fuse their outputs. You cannot fuse a number expressing someone's heart rate with a number expressing the frequency of someone's voice at a given time, but you can fuse the two measurements of two different heart rate sensors to reduce the Mean Squared Error (MSE).

- **Feature-level fusion**. When a system extracts the features expressed by the raw data and combines them, it performs a feature-level fusion. A feature can be some signal's mean during a certain window of time, the name of a posture or a facial expression represented by an array of mark positions, etc.
- **Decision-level fusion**. A system performs a decision-level fusion when it combines the final results produced by each classifier, that is, each emotion detector, these usually being expressed using the six basic emotions (joy, sadness, fear, anger, disgust and surprise) [15] together with percentages (joy 38%, sadness 10%, etc.) indicating how much each emotion is present in the data analysed, although there can be far more affective states identified. On this final level, we retrieve these affective states, together with their corresponding score, and compare them all together to obtain a high-level, richer, and more accurate result. Decision-level fusion is the most commonly used approach for multimodal HCI [2]. The main benefit of this approach over the other ones is that, at this point, all the results are (almost) on the same page, regardless of the format they had when they were first produced.
- **Hybrid fusion**. This approach is actually a combination of feature-level fusion and decision-level fusion. This type of approach arises when we fuse the features of two detectors, classify the results using an automatic classifier and then fuse the result of that classifier with the results produced by another emotion detector (another classifier).
- **Model-level fusion**. This approach is not exactly a data fusion approach since the data are not actually fused but fed into another classifier which has been trained to classify multimodal information.

Since HERA does not implement the emotion detectors, but just requests the results from them, we will always work on the decision-level, using the PAD format that we define below.

2.2 Emotion representation

Another defining feature of emotion recognizers is the way they represent the detected emotions. The definition of emotion itself was already a hot topic of discussion in antiquity, and this discussion is still fully alive today, a fact which has led to the coexistence of 92 emotion theories [34]. Despite this debate about what emotions are, how they are produced or how they are felt by a person, when it comes to representing them using values, there are several approaches which are currently being used [35, 40, 60, 66]:

- *Categorical models*. Also called *discrete emotion representation models,* these representation models use categories in the form of labels to represent emotions. With this type of models, a detected emotion is expressed in terms of the categories we are using. One of the most extended categorical models is the one proposed by Paul Ekman [15], which uses the famous six basic emotions to express an emotion. It is also very common to see the "neutral" category added to this model to represent the absence of emotion. Although these models are easy to understand, it may be difficult to choose the correct categories to accurately represent an emotion.

- *Dimensional models*. In order to alleviate the inflexibility of categorical models, dimensional models were proposed. Instead of having fixed categories of emotions, these models use dimensions which express aspects of an emotion, so what was expressed using a fixed set of categories and a set of percentages is now expressed as a point in a dimensional space (usually 2D or 3D). One of the most widely used dimensional approaches is the PAD (Pleasure, Arousal, Dominance) model, which was proposed in [46]. In this model emotions are expressed in three dimensions, each one in a range of values that usually goes from −100 to 100 or − 1 to 1. These dimensions convey how pleasant (100) or unpleasant (−100) a person's feelings about something are, how aroused (100) or relaxed (−100) a person feels and how dominant (100) or submissive (−100) a person feels while experiencing an emotion. In Fig. 2 we can see a graphic example of this model, in which fixed categories representing emotions are now sectors in a 3D space. This type of models allows us to work with machine learning algorithms in an easier way and also provides us with a form of representation that can become a layer of abstraction for different categorical models, as we will describe later in this paper.
- *Componential models*. This type of models could be considered categorical models as well, since they are built on fixed emotion categories, but they define another dimension of information, defining hierarchies of emotions (Plutchik model, OCC model) [35] or even adaptations for specific fields of research (Hourglass of emotions [5]):

After reviewing the different models, the classic PAD model was chosen as the standard model to represent emotions in our proposed framework. The reason behind this decision is twofold. Firstly, categorical and component models assume the orthogonality of emotions which, in contrast, are overlapping. In comparison, the dimensional model better accommodates the inherent overlapping nature of different emotions. Also, when it comes to

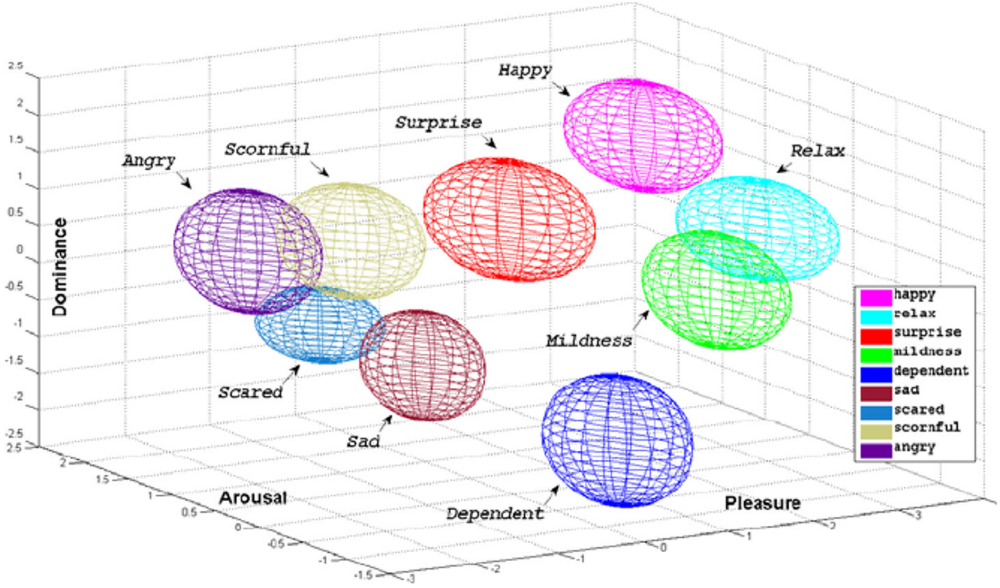


Fig. 2 PAD space [74]

operationalise these models, the dimensional model is easier to implement as well. Secondly, dimensional models help us overcome the heterogeneity of formats commonly used in the industry to represent results. Even when most of the emotion detection services which exist nowadays use a categorical system to represent emotions, they are all different. Hence, developers find themselves facing a problem of heterogeneity of results: they have different services producing results based on the same data, but each service expresses its results in a different way, with a different set of tags, using different ranges of values, what makes it impossible to merge those results directly. To provide a solution to this problem, we have included an aggregation mechanism in the proposal of our framework based on *translating* every result to a standard format. For this purpose, we have chosen PAD, which allows developers to express an emotion as a point in a 3D space. By homogenising all the results produced by each detector, we can aggregate them all together, producing a richer and more solid final result. In following sections, we will see how it was deployed in HERA.

2.3 Research challenges of multimodal emotion recognition systems

Multimodal systems have been the focus of much work in the last few years. New mechanisms for reading emotions have been developed, classification tools have become more powerful and more accessible, and many public databases with multimodal affective data to work with have been created [49, 55]. Over these years, the initial challenges have been met, but new, more complex challenges have arisen. Some of these challenges are explained below [4, 14]:

- *Limitation of the available data to train emotion detectors*. Since emotion detectors are just automatic classifiers, they rely on the datasets they are fed to be able to perform appropriately. For a long time, the shortage of multimodal affective databases limited the amount of works that could be done in this field, even though some multimodal databases have appeared in the last years (e.g., DEAP [36], RECOLA [57], BP4D+ [73]). There is also a problem with the use of prepared datasets, and it is that models trained with these data do not perform well in spontaneous environments [14, 59, 52], although there are proposals to fix this issue [33, 37, 73].
- *No context information being considered***.** MM systems work better when they consider an affective state in the context of previous ones, since this makes it possible to better consider new emotional responses. However, even when there are proposals to consider this information, contextual information is sometimes overlooked [49].
- *Irregular ecosystem*. Unless you have access to a set of trained models to detect emotions in different channels, you are forced to resort to existent, publicly available emotion detectors (assuming you do not have the time to learn how to develop your own model, to gather the media necessary to train it, and to improve it to reach an acceptable degree of accuracy). The ecosystem of the different detectors available over the Internet is quite extensive. For instance, you can find free open-source detectors which are offered via TensorFlow models, services offered on payment of a fee which let one analyse media resources through HTTP requests, etc. The problem here is that each service works in a specific way and has its own advantages, which makes the learning and integration process very slow at the beginning.

Even though these are some of the challenges researchers and developers working in the field of AC have to address, the irregularity of the ecosystem is the most difficult to overcome, especially when developers try to make different services work together. In order to solve this problem, we have developed HERA, a framework to integrate different emotion detection services together that only demands the developer to complete a small template of code in order to integrate a new service into a bigger ecosystem. In Section 3, we will describe how HERA is organised and how it is used.

As it was stated above, the main contribution of this proposal does not rely on the training of an automatic classifier or any other model, but in the development of a framework to integrate very heterogeneous technologies together. This approach does not use machine learning algorithms, as the goal is to merge results based on empirical knowledge.

2.4 Multimodal emotion recognition systems

As we mentioned above, the concept of multimodality has been around for quite a long time now and not just in theory. Multimodal systems have been deployed and tested in fields such as education [1, 13, 20, 27, 29, 32, 68], healthcare [6, 7], marketing [5, 30, 38, 42, 58] or gaming [22, 59].

However, most of the systems developed in these cases were completely ad-hoc, that is, they were developed thinking specifically about the use case they were going to be applied in, there being very little effort on the part of the academic community to develop more abstract systems or frameworks for developing multimodal emotion recognition systems.

Prior to the development of the HERA system, we reviewed the existing literature to study possible frameworks or systems that were created for this purpose. The related works we found are briefly described next. Gonzalez-Sanchez et al. proposed an agent-based software architecture of a generic multimodal framework using design patterns [23], a proposal that they implemented and tested in several real scenarios with different types of inputs [24].

Zheng et al. proposed a multimodal framework specifically to recognise emotion from EEG and eye movement [75].

Maël Fabien’s team, in partnership with the French Employment Agency, developed a multimodal emotion recognition system in the form of a web app (using Flask) which analyses facial, vocal and textual emotions. This system is meant to be used to analyse the face, voice and written text of a person being interviewed for a job offer [17].

Myeongjang Pyeon started the development of a web-based interactive multimodal emotion recognition framework but left it unfinished and undocumented [56].

Alepis and Virbou [1] developed a proposal which is the closest to our proposal we have been able to find, although it is designed specifically for mobile phones and handheld devices.

The W3C Multimodal Interaction Working Group has proposed a framework to develop multimodal systems, that is, systems that allow the users to interact through different interaction channels [65]. Even though multimodality, when it comes to interaction, does not completely correspond to multimodality when it comes to emotion detection, the Multimodal Interaction Framework (MMI) presents a similar architecture to the one proposed in this work. Among the similarities with our approach, this framework (MMI) is also based on different components in charge of receiving input data to produce a result, in this case, multimodal interactions to produce a concrete action on the system. Nevertheless, in the MMI framework, the second level component performs semantic interpretations of the multimodal actions, whereas in our proposal the second level component performs the translation to a common format (PAD) of the different emotion recognition sources (facial, voice, etc.). Regarding emotions, W3C has also proposed a notation to represent them using XML, based on the different emotion models that we presented above [64].

Even though a lot of different implementations of multimodal systems can be found on Github, most of them cover specific use cases instead of giving the structure or tools necessary to create one of those systems. The HERA system aims to fill that gap. In the next section, a description of the system can be found, together with an explanation of the decisions that were made in order to develop it.

3 System description

HERA is a three-level multimodal emotion detection framework designed to manage different emotion detectors in one place, detect emotions in multimedia resources using these recognizers and aggregate results from each emotion detection service. Essentially, HERA (from now on also referred to as the server, the framework, the tool, or simply the system) works like a proxy for other services, these being a third-party emotion recognizer, a device streaming affective or physiological data, etc. Instead of communicating with the services directly, having to spend additional time synchronizing them, performing the authentication process that they may need, discriminating between them depending on the type of resource one wishes to analyse and so on, one can simply tell the system which services one wishes to use and provides it with the settings that those detectors need. When this setting is finished, HERA will simply wait for the analysis requests to arrive, processing them according to the type of media that the request contains and the kind of affective information that the system should seek in this media. HERA will automatically redirect the analysis request to the proxies that can handle these requests, later storing the results that each service produces.

It should be noted that this proposal is not a closed ecosystem with a limited number of services. Our project has been developed following an *Interface-Based Architecture* in order to increase the modularity and maintainability of the system. This means that every detector implements the same interface (see Detector subsection), so adding more detectors (proxies) to the system is a matter of implementing this interface for each of them.

In the next subsections, the HERA ecosystem is described in detail.

3.1 Development approach

This first version of HERA was developed in the form of an API REST using Node.js and the framework Express.js [16]. The development of this framework as an API was based on these principles:

- *Interoperability*. Developing the framework as a module of a specific technology or programming language would have restricted its access to possible users, since researchers or developers working in AC may already be working with a certain technology or framework. If this technology is not compatible with the technology used to develop our tool, or the developers do not know how to use it, they would not be able to use our system, making our tool available only to developers who are familiar with this technology or those that could integrate it with the technology they are already using. In order to facilitate access to HERA and expand our target user group, we developed this proposal as a *web server* (which can also be containerized using Docker), taking advantage of the universality of HTTP communications, which can be implemented with most of the technologies available nowadays, so HERA can be used in any project regardless of the underlying technology.

- *Independency*. When developers are working on an application that involves emotion detection, they have to use valuable time and resources in coding all the aspects related to emotion detection and processing, polluting the codebase with code that does not actually do anything in the app, but which manages aspects that it does not manifest in the final application view. With this approach, i.e., keeping all the emotion processing logic aside, on a separate server, developers just need to take care of capturing the media that they wish to analyse, communicating with HERA and handling the results that the system produces.
- *Easy to extend*. In order to make this framework an appealing multimodal system for developers to use, we need to design it so that it is able to work with many different services. However, we cannot code all the needed functionality to work with every possible device or service in existence. Even if HERA includes some well-known detectors (that is, their corresponding interface implementation), such Face++, which has already been implemented, developers may need to change how the system works with them, or they may need to communicate with a brand new emotion detection service, so this framework needs to be *easy to extend*, using the mechanisms provided by JavaScript in this case so it is simple to add support for new devices.

Based on these three precepts, the HERA system was developed as a web server with Node.js, using the Express framework to build the REST API infrastructure. Thus, our framework is designed to be deployed on a Node server and used as part of a bigger system, as the diagrams in the next figures (Fig. 3 and Fig. 4) show.

In the next subsection, the different artifacts and components of HERA are described in detail.

3.2 Architecture

In Fig. 3, we can see a standard UML deployment diagram which shows our framework in a simulated real context. As explained above, HERA is actually a Web server built with Node.js and ExpressJS, so it can run on a server or in a Docker container, for instance. Once the server is deployed, developers using this framework would communicate with it using HTTP requests, which work with the TCP/IP protocol. In the right-hand part of the diagram, we can see several services as examples of information sources for which HERA could act as a proxy. The first one represents a Generic Emotion Recognition Service, usually offered as an API which requires an access token that the API owners must issue for you previously. It is

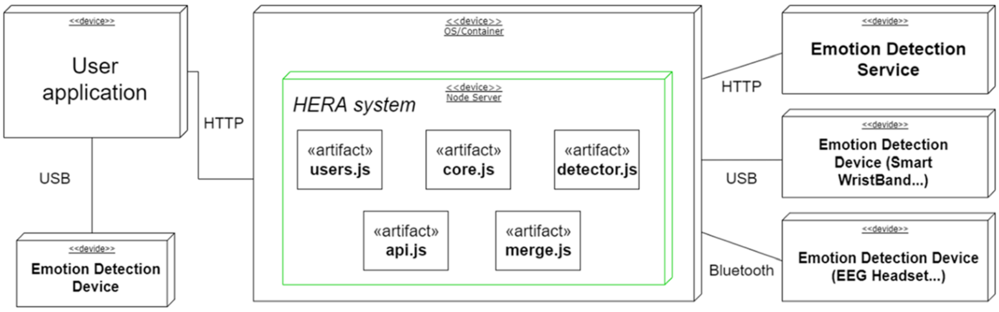


Fig. 3 Architecture - Deployment diagram

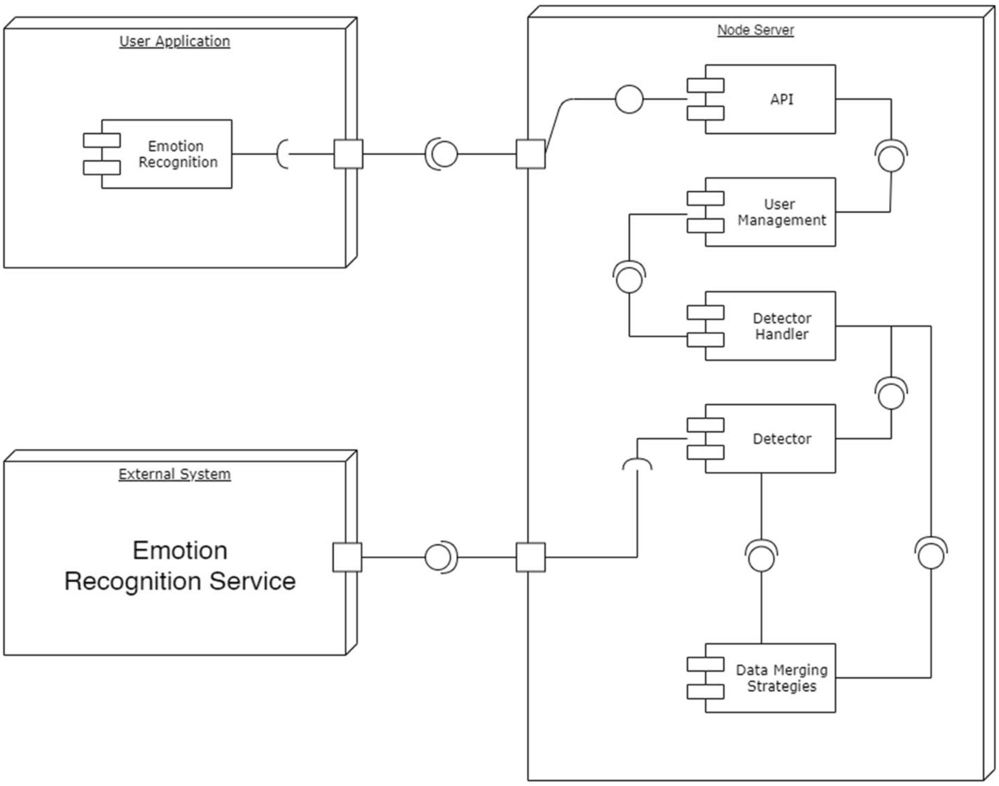


Fig. 4 Architecture - Component diagram

very common to pay for this kind of services to integrate emotion detection in an application, since the companies who makes them offer well-trained emotion recognizers as a service. These services usually work over HTTPS, receiving an item of multimedia which holds affective information (a picture or a video, a sound file, an excerpt from a text, etc.) and returning a set of results in a specific format, usually a categorical approach (the emotion detected is expressed in terms of Ekman's six basic emotions [15]). However, this service could also be a local server providing an emotion detection service implemented by the developers themselves using the *scikit-learn* package, for instance. The other three nodes represent physical devices that could be attached to the device running our server or to the User application directly. These physical devices could communicate their results via HTTP (through an intermediary system) or directly to HERA, using more direct protocols such as USB or Bluetooth. In the case of devices measuring physiological variables, the results would be *rawer*, requiring some processing on the server before they can be translated into a more understandable format, but we will look at this topic in the next subsection.

Returning to Fig. 3, we can see that the system is made up of five main artifacts:

- *api.js*. This file holds all the API-related logic. Even though there are a couple more files related to the Express framework and the API logic, the declaration of the different endpoints that the developers reach from their external applications can be found in *api.js*.
- *users.js*. In order to be able to keep track of requests coming from different applications (or from the same application but for different purposes), HERA implements sessions using the *users.js* artifact. This artifact contains logic to create new users, which are actually

sessions, using cookies, to add information (detectors) to these sessions, to refresh sessions and to delete them when a certain timeout is reached.

- *core.js*. This artifact contains the *core* functionality of the API, that is, the management of the different detectors, all of them organized in different groups depending on the affective channel in which they can detect affective information and the orchestration of the subsequent data fusion (after performing the analysis of media resources).
- *detector.js*. This artifact contains the key building block of the API: an abstraction for a device or a system providing affective information. For each emotion information provider, whether it be a physical device, a paid API offered by a third party or an automatic classifier trained to identify emotions running locally, HERA creates a *Detector* object, which will act as a proxy for this service, handling the requests that usually would be sent directly to it. Following the principle of keeping the system modular, each detector implements part of its functionality as an external module that is imported when the Detector object is created. We will review this in Subsection 3 of this section.
- *merge.js*. This artifact contains the different strategies that can be used to fuse the affective data. These strategies receive a list of PAD triplets, which come from the results produced by the different detectors and return a single PAD triplet. In addition, since the strategies also follow a *modular* approach, they can be defined by following another format or adding more elements (or properties) to the final result, creating new information based on that which we already have, and this is one of the main assets of HERA.

In Fig. 4 we can see a component diagram showing how the different artifacts work together to fulfil the users' requests:

- The *api.js* artifact from Fig. 4, here represented by the API component, exposes the API endpoints that users reach using HTTP requests. This API has been developed using Express.js, and the different endpoints can be reached using GET and POST requests. In the following subsections, we will specify how many endpoints HERA offers and what information they need to operate.
- As stated above, the system uses sessions to tell incoming requests apart and to store detector handlers, so the API component needs the functionality exposed by the *User Management* component, which corresponds with the *users.js* artifact.
- Likewise, the User Management component uses the logic offered by the *Detector Handler* component (implemented in the core.js artifact) to attend to the users' requests. This component contains logic to add new detectors to the current session, to organize them into groups depending on the input they can handle, to forward requests to the corresponding group and to fuse the data available in every detector.
- In order to use a detector in this framework, developers must have previously implemented a set of methods so this detector can be handled by the system (although HERA already includes several Detectors that are implemented in its repository). These methods constitute the *Detector* interface. In the following section we will see this interface in detail.
- Finally, both detectors and the Detector Handler of a session need access to the data fusion strategies, which are managed by the *Data Merging Strategies* component. Each detector use strategies to merge their own results into a single value (a PAD triplet), while the Detector Handler uses strategies at two different levels: first, it fuses the results coming from each category of detector, and afterwards it fuses together the results produced in this previous first fusion. We will see how this process works in the following subsection.

In the next subsections we will review how the data fusion is performed, how the detectors are integrated in the system and what would be the regular workflow of the framework.

3.3 The detector interface

As we mentioned at the beginning of Section 3, in order to be an effective tool, HERA must be able to work with all types of affective information providers, whether it be an API receiving requests over HTTPS and returning JSON objects with a certain content, a physical device streaming an array of data via a USB port, or a device communicating with the system through Bluetooth packages. Not only must the system deal with data in very different forms, but also with the idiosyncrasies imposed by each detector (authentication processes, gathering of data, etc.). In order to deal with this range of possibilities, HERA uses the D*etector* class as an abstraction layer, treating every different detector as a Detector object. Every detector has an *id* (which is usually its name), the *kind of content* in which it can find affective information (face, voice, body, etc.), a list of the media formats it supports (images, video, audio tracks, etc.), a *delay* property which indicates the average response time of the detector, the *address* where the detector delivers its service (if any), and two *lists* storing the results produced, one with the original results and one with the corresponding PAD translations. In an object-oriented (OO) programming language, Detector would be an abstract class and every specific detector would be a new class which would extend it, adding the functionality that this new detector could need. However, since basic JavaScript does not support abstract classes directly, HERA implements it using modules: every detector is implemented as a *module which must export three key functions*, which are associated with a Detector object in runtime. These three functions are:

- *initialize.* This method takes care of any initialization tasks that your detector needs. For instance, if this detector is a proxy for a third-party service offered via Internet, developers may need to obtain an *auth* token first. If this detector is a proxy for a bluetooth wristband, developers may need to put their discovery and connection code in this method. If developers do not need any initialization, they should just return a resolved promise (*Promise.resolve*).
- *extractEmotions.* This is the main method of every detector, the one in charge of performing the actual emotion detection, whether this be forwarding a media resource to an API over the internet, reading RAW data from a sensor, etc. This method receives three parameters, namely *context*, *media* and *callback*. *Context* is the environment from which the method is called, usually a Detector object. This is done so that these methods can be specified in a separate file, but results can be stored correctly in the corresponding Detector object. *Media* is a path to the file which holds the affective information, if there is any. If there is no file (maybe the detector reads some RAW data from a certain port or socket in this method), this parameter will stay unused for the sake of the aforementioned interface-based programming paradigm. The final parameter, namely *callback*, is an optional callback to handle the retrieved data, in case further manipulations are needed.
- *translateToPAD.* This method is one of the keys of this proposal, since it is the part of HERA that allows us to integrate heterogeneous emotion detectors. In order to actually be able to treat all the detectors in the same way, we need to translate all the results they produce to the same language. The chosen language for this task is the PAD format, which we introduced in Section 2. This method must implement the transformation of the

incoming results, so the affective results (whatever their format is) received by a detector are translated into a triplet of three numbers, each one being a value between −1 and 1, which stands for how negative or positive the expressed emotion is, how relaxed or excited the person is, and how passive or dominant that person feels while experiencing that emotion, respectively. In Fig. 5, we can see an example of the type of transformation this method does.

Regarding the organization of these different modules, it is highly recommended to keep them organised in folders according to *categories* (a `face` folder for modules of detectors which support facial-expression-based emotion detectors, a `voice` folder for voice-based emotion detectors, etc.), although this is not actually necessary since the path to a given detector's functions is fully specified in the request for setting up the different detectors.

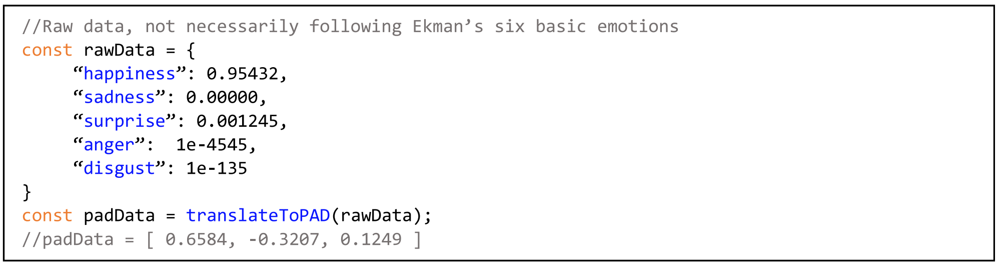
In the next subsections we will see how this uniformized data is integrated and how these methods are associated with their corresponding detector.

3.4 Data fusion

In Section 2 we introduced one of the key aspects of a multimodal system, namely the fusion of different types of data coming from different sources of affective information. Since multimodal systems gather information with different formats coming from different sources, they need to perform a transformation and/or an integration task of this data in order to produce a coherent final result. Therefore, the first thing to do when a new result is produced is to translate it into its corresponding PAD value. Every detector keeps a list of all the results they have produced, so the data fusing challenge is *to turn a set of lists of PAD results into a single result while considering the context of all the detectors*, as some of these results may have been produced by *different detectors of the same type* analysing the same media file (e.g., two face-based emotion detectors analysing the same picture), by different detectors analysing media resources with *different formats* (e.g., one result produced by a face-based emotion detector and one result produced by a voice-based emotion detector), by detectors with a very different rate of result production (e.g., one detector has produced 3 results and another detector has produced 1.000 results), etc.

We have designed a fusion framework which fuses data three times, one per level:

- The first level is made up of the specific detectors that have been integrated into HERA, that is, emotion detection services which had had their Detector interface implemented. In other words, we say that an emotion detection service has been integrated into HERA



```
//Raw data, not necessarily following Ekman's six basic emotions
const rawData = {
     "happiness": 0.95432,
     "sadness": 0.00000,
     "surprise": 0.001245,
     "anger":  1e-4545,
     "disgust": 1e-135
}
const padData = translateToPAD(rawData);
//padData = [ 0.6584, -0.3207, 0.1249 ]
```

Fig. 5 Example of translateToPAD

when we have created a file which contains the three methods mentioned in the previous section to communicate with said detection service and to translate its responses. The fusion in this level is done by telling every detector to aggregate its results from the current session into one result. At this point, each detector has turned its list of results into a single PAD triplet.

- The second level is made up of the *categories of detectors* detecting emotions in the same modality. HERA organises detectors by placing those which can detect emotions in the same modality under the same group, this way, when a request of emotion detection over this category arrives, HERA forwards this request to each detector capable of fulfilling it. The fusion at this level is done by aggregating the results produced by the fusion in the previous level together, always grouping these results by category. At this point, HERA has a PAD triplet per category. Since the detectors of each category have all produced a result for each request received, aggregating all the results of the same category into one helps to decrease small discrepancies.
- In the third level of this data fusion framework, HERA aggregates the output of the previous level into a final single PAD triplet. The aggregation strategy at this level might be more complex than the one used in the previous levels, since now we might have to consider different weights for the different categories, metadata attached to the PAD triplet produced by each category, etc.

Figure 6 shows a sequence diagram that models the data fusion process throughout the whole system:

- The user sends a merging data request to HERA through the *results* endpoint (Fig. 7). As it is stated in the documentation, this request demands three parameters: a list of channels

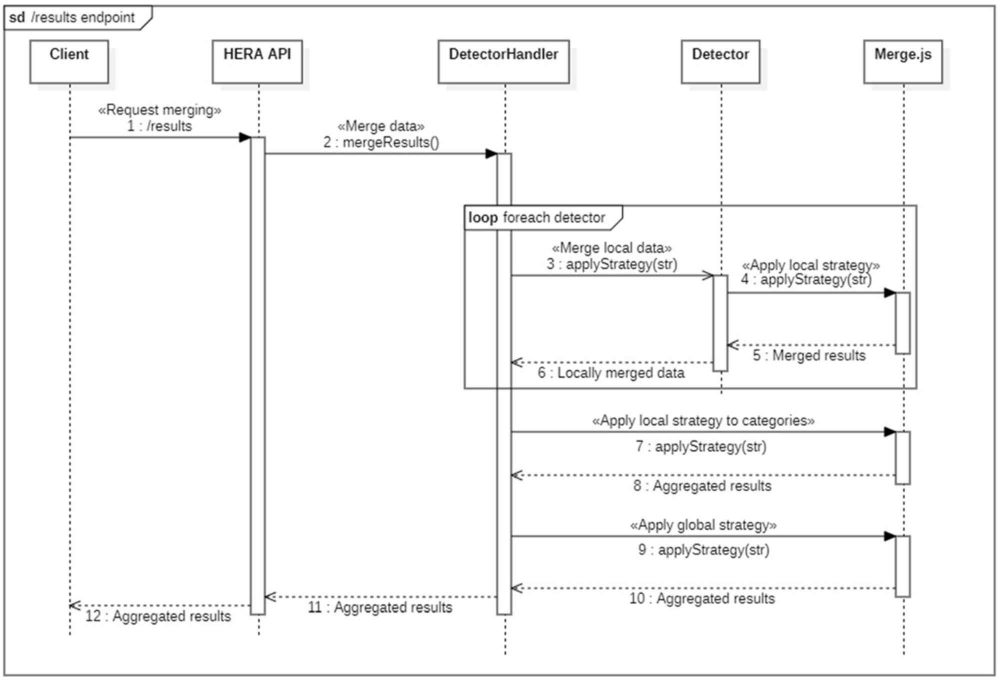
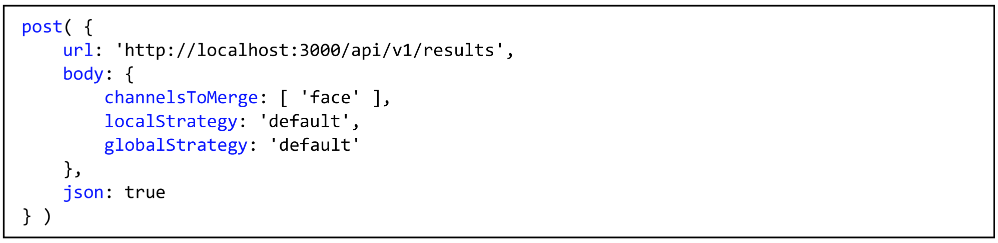


Fig. 6 Fusion strategy diagram

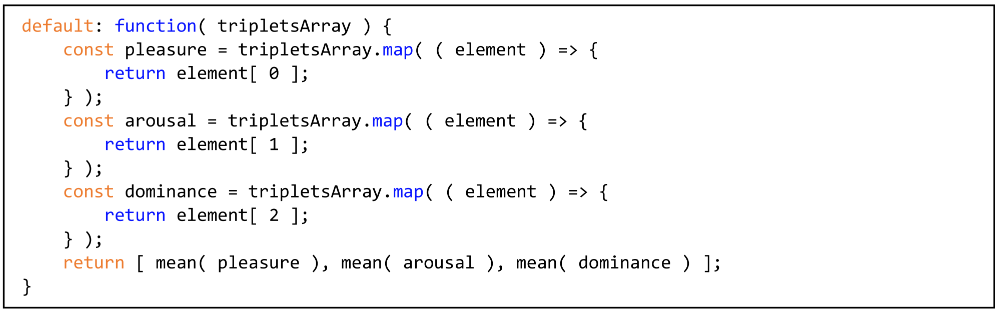


```
post( {
    url: 'http://localhost:3000/api/v1/results',
    body: {
        channelsToMerge: [ 'face' ],
        localStrategy: 'default',
        globalStrategy: 'default'
    },
    json: true
} )
```

Fig. 7 /results request endpoint

that are going to be merged, the name of the strategy that the API has to use to aggregate the local data and the name of the strategy that the API has to use to aggregate the data of detectors from different categories. In Fig. 7, we can see a request to this endpoint using JavaScript and the *request* package. In this snippet, we are asking HERA to fuse all the data from the "face" and "voice" channels. In our tool, a channel is a group of detectors which can extract emotions from a certain modality. For instance, the channel "face" a list of detectors which can read emotions from the face of a person (facial expression). The channels' names can be fully customized by the users of HERA.

- When HERA receives a "/results" request, it forwards the request to the Detector Handler component of the API, which keeps the detectors organized in *categories*. Since each detector on each category holds a list of results in PAD format, HERA must combine all these results into one. To do so, HERA will request every Detector object to aggregate its own results using the local strategy *localStrategy* that was attached to the request.
- Each Detector object keeps two lists of the results produced in the current session: one list contains the results in its raw form, while the other one contains the PAD version of those same results. When the *mergeResult* method is invoked, the Detector calls the *applyStrategy* function from the *merge.js* module, which applies the fusion strategy specified in the aforementioned request to a list of PAD results. The merge.js module has all the implemented strategies organised in a dictionary. In the context of HERA, a *strategy* is a function which receives a list of triplets, each triplet representing an emotion in a PAD space, and returns a single triplet. In Fig. 8, we can see an example of a strategy, which performs an average operation over the different coordinates of the PAD values of the list. Since HERA is implemented in JavaScript, we can also take advantage of the capabilities of the language to add extra properties and metadata to this final result.
- Once every detector has aggregated its data using the *localStrategy* strategy, they return this single PAD data to the Detector Handler object. At this point, HERA has a PAD triplet per Detector, being these detectors organised in categories. The next step is to aggregate



```
default: function( tripletsArray ) {
    const pleasure = tripletsArray.map( ( element ) => {
        return element[ 0 ];
    } );
    const arousal = tripletsArray.map( ( element ) => {
        return element[ 1 ];
    } );
    const dominance = tripletsArray.map( ( element ) => {
        return element[ 2 ];
    } );
    return [ mean( pleasure ), mean( arousal ), mean( dominance ) ];
}
```

Fig. 8 Strategy example

the PAD triplets from the same category into one. To do so, HERA will repeat the operation from the previous step: for each category, the Detector Handler object will invoke the *applyStrategy* method from merge.js, passing a list with as much results as detectors objects there are in that said category. Because of the similarity of results, the same *localStrategy* strategy from the previous step is used.

- Once HERA has obtained a PAD triplet per category, the Detector Handler will call the *applyStrategy* function again with these triplets as arguments, using the *globalStrategy* strategy this time. Usually, the *localStrategy* strategy will be an easier operation since the data to aggregate is much more similar, while the *globalStrategy* strategy will probably consider the context of the each triplet, possible metadata added to the triplets, etc., although the same strategy could be used in the different steps of the process.
- Using the *globalStrategy* strategy, a single PAD triplet, which could possibly hold additional metadata, is produced. This final result is sent back to the user, finishing the merging workflow.

3.5 System workflow

After having reviewed the different parts of the proposed framework, now we will see how all these pieces work together (Fig. 9). Once the server is running (deployed on a Node server and attending to requests in a specific port), the workflow is as follows:

- *Requesting an access token*. As we mentioned above, HERA uses sessions to distinguish where the requests are coming from and/or to group different sets of detectors. These sessions are implemented using cookies, which store an access token that the server provides us with. This access token must be included in any request directed to the server in the future, otherwise we will not be able to access its functionality, and it will warn us about the absence of an access token. In order to obtain our access token, we must send a request to the API's root (endpoint "`/`"). In response, the system will create our session on the server and send us in return our unique access token stored in the *userId* cookie. This id allows us to communicate with the rest of the endpoints. Trying to communicate with these endpoints without including this unique id in the request will return an error ("`Session wasn't initialized. Send request to "/" first`"). This id is linked to a `user` object which will store the detectors' proxies, the data returned by these, etc.
- *Setting up the server*. Once we have our access token, we can communicate with HERA so that it creates a proxy for each detector we wish to use in the current session. To do so, we must send a request to the "`/init`" endpoint, attaching to said request the information about the detectors we wish to use. This endpoint will link a `DetectorHandler` object to our user session and will create a `Detector` object for each detector specified in the request (this setting information can be included directly in the request or stored in a file on the server if we are always using the same detectors, in which case we must specify the route of this file in our request). Our system also supports a benchmarking process for each detector: if we provide each category of detectors with a folder containing resources that those detectors can analyse, setting up a detector at this endpoint will start a benchmarking task, in which all the resources from the corresponding folder will be analysed using this detector. From this task an average response time is calculated so that we can later filter detectors out depending on their response time.

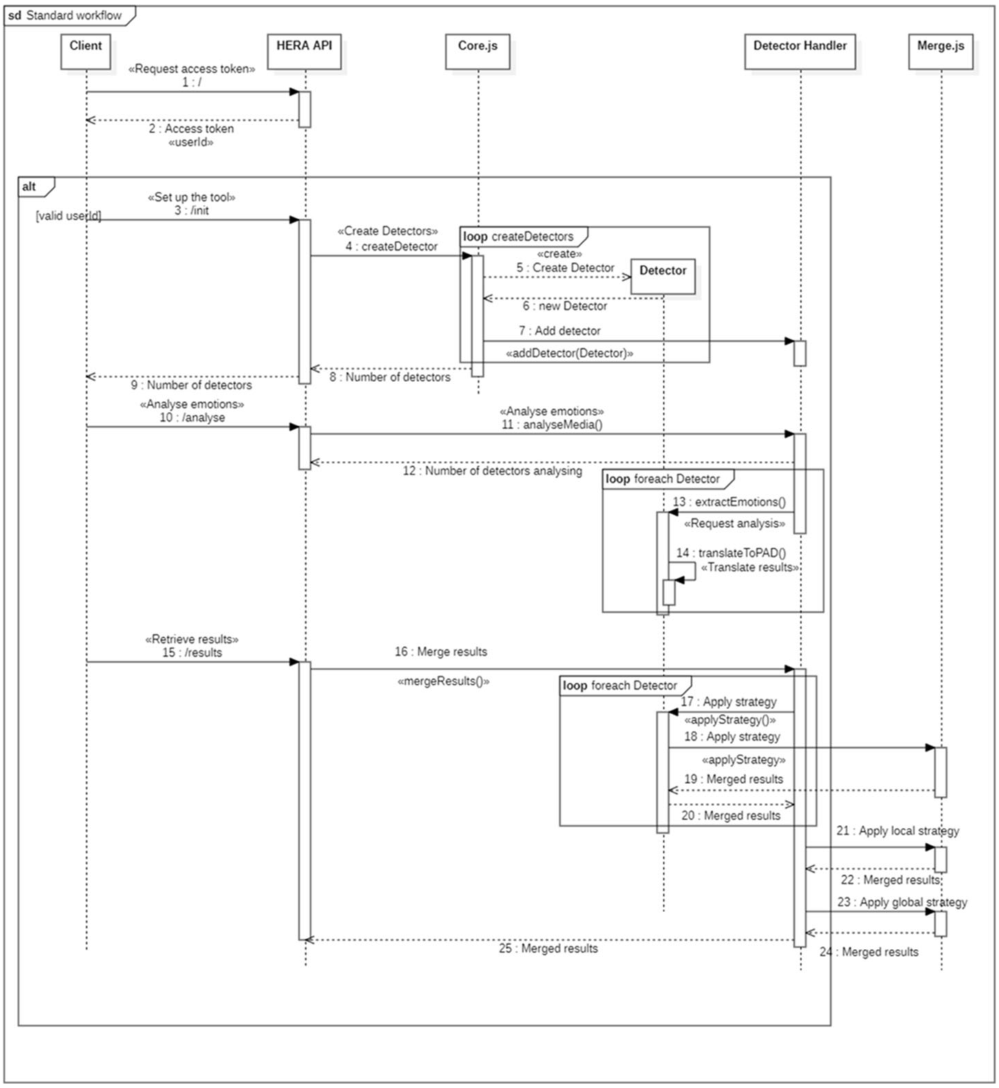


Fig. 9 System workflow

- *Requesting a media analysis looking for emotions*. After calling the previous endpoint, HERA is ready to handle our emotion analysis requests. In order to request an analysis, we must reach the "`/analyse`" endpoint, sending to it the resource we wish to analyse, if any (the server may need to pick this resource itself from a socket or a USB port), the type of resource it is and how the affective information is coded, that is, whether it has to analyse the resource by looking for faces, for voices, etc. HERA uses this information to forward the requests to the detectors capable of taking this request, all of them organized in categories under our `DetectorHandler` object. This endpoint uses the `extractEmotion` function of each detector to retrieve the corresponding results, and the `translateToPAD` functions to store the results. The response to this endpoint does not contain the results, since they must be requested through the next endpoint.

- *Asking for results*. Finally, once the detectors have produced the required results, developers would poll on the HERA server to ask for the integration of the produced results using the "`/results`" endpoint. In order to communicate with this endpoint, we have to indicate what data merging strategies HERA should use to aggregate the results and what channels should be aggregated together. In this way, this endpoint can be used to retrieve the RAW results for a single channel, a PAD triplet with additional properties produced by aggregating all the existent results, etc.

The framework also includes another endpoint which developers can use to filter detectors out once they have been created. Thus, developers have an additional tool to control their detectors.

- *Filtering detectors*. After creating the detectors, we may wish, for instance, to filter out detectors which are too slow, or maybe we have lost one of our affective channels and we wish to remove that category of emotion detectors. To do so, we can send a request to the "`/setup`" endpoint, indicating which categories we are actually going to need or a threshold that the average response time of each detector must satisfy (to remove detectors which take longer than that threshold to complete a request). As per the rest of the API, this endpoint can be modified to add options and filters.

4 Framework evaluation

An evaluation of the system was conducted with seven software developers, all of them with previous experience in the development of applications and in the use of APIs. In this section, the different aspects of the assessment are presented.

HERA was developed under the hypothesis that separating the emotion processing logic from the main application, and providing a modular data fusion framework would help developers to integrate emotion detection in their applications more quickly and more easily, and also give them the possibility of further exploiting the affective information they could gather. In this evaluation we seek to prove this hypothesis.

4.1 Participants and context

For this experiment we recruited seven software developers aged between 22 and 35 years old, all of them with experience in the development of server-side applications and in the use of third-party APIs. Any of them had previous experience with emotion-related technologies, although two of them were familiar with AC concepts. We decided to recruit developers with no previous knowledge on emotion detection technologies so that they could focus on the process of integrating technologies together from a beginner's perspective. According to Jakob Nielsen, even five participants are sufficient to detect 85% of the usability problems in a system [48]. Although seven participants may seem a small number, as the main goal was to assess the validity of the system by specialists in the field of software development, we considered more important to have representative participants than to recruit many people with similar skills repeating the same tasks.

The setup of the experiment consisted of two devices. The first device was a web server deployed on the Amazon Web Services (AWS) platform, which is accessible at all times. This platform was responsible for keeping a HERA server instance running. Evaluation tasks involving HTTP requests to the system were carried out by directing those requests to this server. The second computer was the computer used by each developer participating in this test to perform the corresponding evaluation tasks. Prior to the evaluation, each developer was asked what equipment they would rather use, and all of them decided to use their personal computer, even though they were also offered a computer which had all the software they might need to complete the evaluation tasks. The HERA server was launched using Node.js v12.14.0 and Express v4.16.0, and the evaluation tasks were performed using Node.js v12.14.0 and two different Python versions (2.7 and 3.7).

4.2 Evaluation metrics

After reviewing previous studies on API usability assessment [2, 10, 26, 62, 67], especially the API evaluation framework used by The Visual Studio Usability group at Microsoft [9], we decided to adopt a combination of the Twelve Cognitive Dimensions Questionnaire and the Computer System Usability Questionnaire (CSUQ) [3, 25, 41]. The Cognitive Dimensions framework was proposed in 1989 as a framework to evaluate the different aspects (dimensions) of a programming language, but in the last few years it has been applied to specific products, such as APIs, for instance. These dimensions are the following: abstraction level, learning style, working framework, work-step unit, progressive evaluation, premature commitment, penetrability, API elaboration, API viscosity, consistency, role expressiveness and domain correspondence. To apply this framework, we prepared a questionnaire with twelve questions, one per dimension, so that participants could manifest their impressions of the API regarding each dimension. For example, does the API expose all its functionality, or does it expose only certain abstractions of different functionalities? How easy is to make a change? Does the API work in a consistent way? This is a sample of the kind of questions that the users had to answer. In line with [9], we gave this questionnaire to the participants before and after performing the evaluation, to contrast their expectations of how the API should work against their impressions of how it actually works. Regarding the CSUQ, this is a questionnaire of 19 questions, all of them answered with a 7-point Likert scale, about the usability of the product being tested [39]. Altogether, users complete three questionnaires: one to check their expectations, one to gather their impressions of the tool (using the cognitive dimensions framework), and finally one to measure the usability of the tool.

4.3 Experimental design

After considering previous works on the assessment of APIs [2, 10, 26, 62], we designed the following evaluation framework:

(i) *Research design*. The experiment was designed as a within-subjects experiment, so users could appreciate the process of using the HERA system against the use case of not having it, that is, having to communicate with several emotion detectors manually, sync the return of results and the posterior fusion of data. Hence, each participant is put through three programming exercises which are essentially identical. This exercise consists of a set of tasks whose final goal is to communicate with two emotion detection services, so they analyse a media file looking for emotional responses. The participants wrote three programs to: (1) communicate with two different emotion detection services, namely Face++ and DummyDetector; (2) request an emotion analysis of a given media file, a picture in this case; and (3) fuse the different results of each detector into a single metric. For the first exercise, the participants chose their preferred programming language to communicate with these emotion recognition services. For the second exercise, they used the same language to write requests as if they were going to be sent to our server, while for the third exercise they had to use a different programming language to actually communicate with our server, although all of them used JavaScript or Python. The second exercise serves as a training process in the use of HERA, in which users learn how to use it, which involves communicating with the different endpoints, configuring the proxies on the server, and asking for an analysis and the corresponding results. In the final exercise, the users recovered the requests they wrote in the second exercise and adapted them to be sent to the server that was actually deployed, so that it would be able to act as a proxy for the Face++ and DummyDetector services.

(ii) *Intervention*. An instance of the server was launched using the Amazon Web Services Elastic Beanstalk service, which allows us to easily deploy a web application or infrastructure. As was stated in the instructions script that was given to the participants, interaction with the system would be carried out through HTTP requests directed to the AWS server. The participants were also offered a computer prepared with all the software needed to complete the different programming tasks, but they were also offered the possibility of using their own laptops if they were more comfortable programming on their own computers.

The whole evaluation process was divided into four parts.

(i) *Introduction to the Test*. The participants were introduced to the HERA System and to the context in which it exists, covering such things as the existence of different emotion recognition systems, how every system is different, how there is no norm for how things should be done, etc. In this first phase, they were given a document with instructions on what they had to do during the assessment, and information they would need in the process. After reading this script, they would complete the Twelve Cognitive Dimensions Questionnaire, answering the questions by thinking *only* in terms of the expectations they had about how the HERA System *should* work.

(ii) *First part of the Test*. In this part of the test, the participants had to communicate with Face++ and DummyDetector to ask for the emotions presented in a picture of a face. For this task, they used the programming language they preferred in order to communicate with the different servers, to store the results and to combine them in some way (they were advised to take the mean of the results representing the emotion "happiness" according to each different service).

(iii) *Training with HERA*. Once they had finished the first part of the evaluation, they performed a *simulation* using HERA. Using the programming language of their preference again and the documentation from the evaluation script and from the framework's repository [18], they had to write a structure for a set of generic requests aimed at each endpoint of the API.

(iv) *Second part of the Test*. In this second part, the participants communicate with the HERA system in order to ask *it* to carry out the previous task for them. The participants communicate with an instance of the system deployed on AWS to authenticate themselves, to set up the different proxies for each detector (Face++ and DummyDetector), to ask for an emotion recognition analysis on a media file (a picture) and a subsequent

Table 1 Participants data

Participant	Total time	Time first phase	Time training phase	Time second phase	Errors	Assistances
Participant #1	0:52:28	0:25:55	0:21:20	0:05:13	0	0
Participant #2	0:59:05	0:27:32	0:23:51	0:07:41	1	0
Participant #3	1:21:11	0:39:46	0:23:24	0:18:01	1	4
Participant #4	0:50:50	0:20:36	0:20:55	0:09:19	1	0
Participant #5	0:50:33	0:14:42	0:30:30	0:05:21	0	5
Participant #6	0:34:52	0:15:34	0:14:06	0:05:12	1	2
Participant #7	0:37:55	0:14:24	0:17:27	0:06:04	0	2

fusion of the results. This second part is done with a programming language that is different from the one used in the previous exercise. The goal of this is to make sure that the users focus on the behaviour of the framework, and that they are not distracted by the idiosyncrasies of a specific programming language.

(v) *Completing usability questionnaires*. After completing the three programming exercises, the participants completed the Twelve Cognitive Dimension Questionnaire again, this time answering the questions by considering their actual impressions of the tool, and the Computer System Usability Questionnaire. In this last questionnaire, a set of demographic questions were included.

The data collected during the evaluation were subsequently analysed, and the outcomes are described below.

4.4 Evaluation outcomes and discussion

During the evaluation, we measured the time it took the participants to finish each part, together with the errors they made (regarding the use of HERA, not the libraries they need to make the HTTP requests), and the times they needed help from the evaluator. In Table 1 we can see the time each participant needed to finish each part, the errors they made and the times they needed assistance (the participants' data have been anonymized).

In Fig. 10 we can see the average time the participants needed to finish each part. This parameter shows a dramatic reduction in the time needed to finish a task, thus validating part of

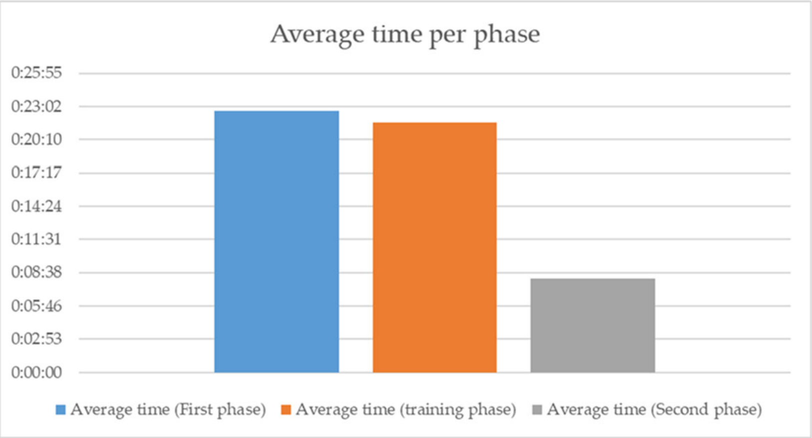


Fig. 10 Average time spent by phase

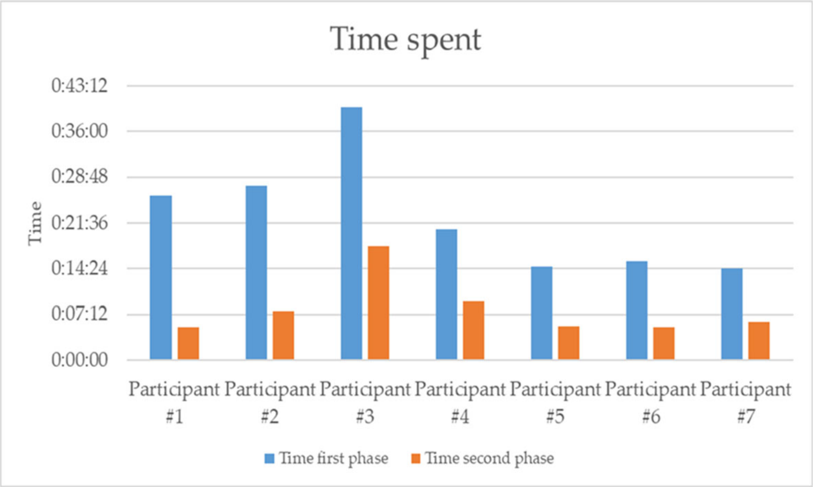


Fig. 11 Time spent per phase and user

our initial hypothesis, namely that after a previous training phase, which is not very expensive in terms of time, the use of HERA becomes very straightforward. Also, we must bear in mind that, since the evaluation is highly scripted, the participants save a lot of time in the first phase that is not reflected in the data gathered. The tasks of studying how to integrate emotion recognition in the system, how to gather the media, and analysing the different response formats of each detector (without having an evaluator helping you to dissect an HTTP response or an intricate JSON object) are all highly time-consuming, and they are *not* reflected in the first column of the figure, even though this phase already took the longest on average.

In Fig. 11 we can see a more detailed view of the data from Table 1 (omitting the time from the training phase). In this figure, each pair of columns represents the time it took a participant to finish the first and second phase. It is clear that the first phase took much longer to complete (around 20 minutes), in comparison with the second one (between 5 and 7 minutes). Due to the characteristics of this sample (small groups, little data dispersion), we carried out a Mann–Whitney U test to mathematically assess this time difference between the two phases, which led us to state that the two sets of data are significantly different with an alpha value of 0.05 (U = 3.0, p = 0.00364).

Finally, we can see the results gathered by the different questionnaires in Fig. 12 and Fig. 13. In Fig. 12 we can see the average response for each question. Since this questionnaire was composed by placing the positive value of the answer as the highest value (7 in this case), the higher the average response, the more positive the answer is, and the higher the total score, the higher the user satisfaction is. In this case, the average response for almost all the questions is above 5, with an average standard deviation of 0.9993, excepting the ninth question.

This question, which asks about the clarity of error messages returned by the API, has a standard deviation of 1.66 (a high value for such a narrow set of options). This difference of opinions was actually explained by the participants who gave a low value as an answer in later interviews.

The two users who answered that question with the lowest values had several issues with the HTTP library that they were using (bad URLs, wrong way of attaching data to the request, etc.), which gave them message errors they did not fully understand, due to a lack of knowledge of that particular library. Since they associated these vague errors with our server

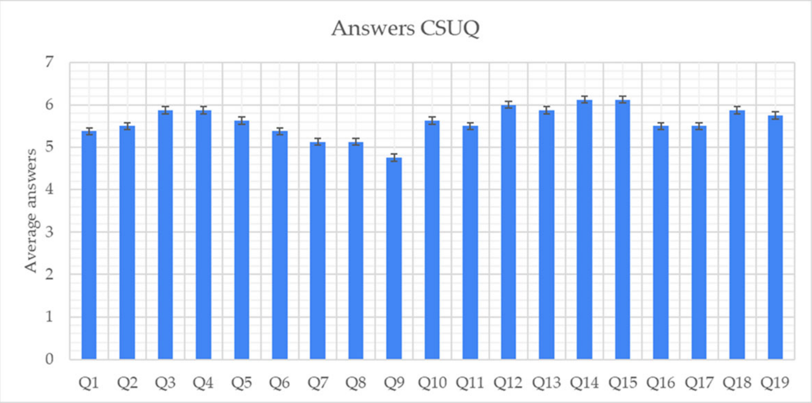


Fig. 12 Average answers of the Computer System Usability Questionnaire

(when the server was not even receiving the request), they gave a low score in that field. The rest of the questions received, on average, good scores. The total average score of this questionnaire is 106.5 out of a total of 133, which equals to 8.01 out of 10. This mark, along with the small deviation for each answer, can be interpreted as an agreement from the participants regarding the high usability of the system.

Finally, in Fig. 13, we can see the two spider diagrams we obtained by plotting the answers given to each dimension of the Twelve Cognitive Dimensions Questionnaire. By following the approach in [9], we surveyed the participants before and after, in order to be able to contrast their expectations and their impressions of the application. On average, we can see how our framework met most of their expectations, even improving upon what they were expecting, especially in the dimensions of Learning style, Premature Commitment, API Viscosity, Role Expressivity and Domain Correspondence.

With these results, we can conclude that the recruited developers found the tool useful to integrate technologies they were not familiar with. It is important to remark that this was the goal of this assessment. Since there are no other tools like HERA on literature to reduce the slope of the learning curve when it comes to integrate different emotion detectors and develop multimodal detectors, developers usually find it difficult to start developing this kind of detectors. This may lead to the development of defective multimodal systems or even to the abandonment of a project, in favour of a simpler unimodal detector. Although we reviewed some projects about multimodal detectors in section 2.4, these were designed ad-hoc for specific purposes or platforms, so the effort of trying to adapt these to a new scenario might be

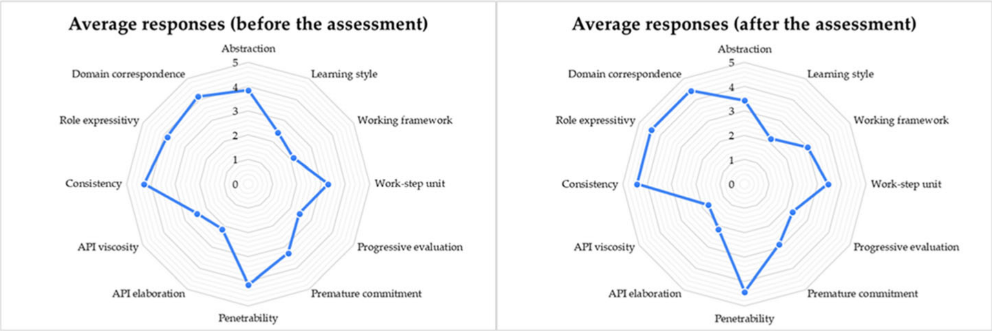


Fig. 13 Average responses of Twelve Cognitive Dimensions Questionnaire

even bigger than the effort of researching multimodality properly and develop a solution from the ground up. HERA is an effort to assist those developers in the elaboration of multimodal systems. However, further assessment with more archetypes of developers is still needed to assess the tool.

Regarding the application of HERA, this framework has been designed as a separate application that can be launched anywhere thanks to the underlying technology that we have used in its development. This approach just adds another entity to any existent architecture, leaving the mainframe of this existent application cleaner than it would be if all the emotion-related logic had to be written from scratch in the system. Some of the scenarios that would benefit from the use of HERA are the following:

- *Applications with educational purposes*. The idea of HERA actually spawned during the development of multimedia applications with education purposes [20, 21], when the development of said applications exposed the limitations and difficulties of building multimodal systems. In the context of these previous works, a tool deployed in the backend of the application would have helped us to develop the system, even if it had needed some customization.
- *Applications handling multimedia content*. Part of HERA needs to be capable of handling different types of multimedia input, being it in the form of a stream, a file, etc. Even if this first version has been more focused on the fusion of data, the proper management of multimedia information can be a valuable asset of HERA on its own.
- *Applications involving business processes*. Some of the activities that companies carry out, like costumer attention, can harvest great benefits from opinion mining and other forms of emotion detection. A system like HERA can help develop a more tailored solution to analyse the input of clients.

5 Conclusions

In this paper we have reviewed HERA, a framework developed in JavaScript to support the creation of multimodal emotion detectors and their subsequent integration with other applications, and the evaluation process that we carried out to test the first version. HERA has been developed while following the principles of modularity, independency and extendibility, since it is offered as an open-source tool that other developers are encouraged to adopt and help grow. With this tool, we strive to fill a gap in the existent literature, bringing together theory and implementation, and supporting our proposal in a standard format, namely the PAD classification system. The system proposed in this paper is designed to offer an easy way of including emotion recognition in an existing system in a straight-forward way, and reducing the usual workload involved in this task. Integrating our framework in another application's infrastructure is a matter of launching a Node server running the framework and sending HTTP requests to it, with this having only a minor impact on the app source code. Since the framework has been developed over Express and is open source, a new web application could be developed by even using the framework source code as a starting point. Although using this framework is not completely immediate, since it requires a previous setup and user training phase, we have proved empirically that the benefits of using this framework, in terms of time and effort, are worth this preparation phase. We tested the tool with real users, who used it to develop a small demo, which revealed some of HERA's strong and weak points.

One of the current challenges is to go on developing HERA so that it can support more forms of interactions and new detectors. Even though one of its key aspects is that it is easy to extend, the more features it has from the beginning, the more likely it is that new developers use it, thus enlarging the community of HERA's users, which will help us to find bugs and make HERA bigger, safer and more solid. In order to tackle this challenge, we will review existent mainstream emotion detection services to integrate them into HERA. This extension process will also include popular handheld and wearable devices, which will contribute to extend the communication capabilities of HERA.

This first version of the framework is specially focused on the fusion of data, but HERA can still be improved to increase its capabilities regarding the management of media input, so that developers can stream multimedia data to HERA and request analysis over these continuous streams, instead of making punctual requests each time.

When it comes to the fusion of data, HERA can greatly benefit from having a wide variety of algorithms to aggregate data all together. We will consider different usages of this tool to develop algorithms to cover the different needs of future developers.

Lastly, the management of results within HERA can open the door to machine learning, that could be applied once we have gathered enough results, previously validated by experts. This could help to triple validate the aggregated data and/or to increase even more the richness of the data.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Availability of data and material Any additional data or material could be provided by the authors if needed.

Code availability The code of this framework is available at [18].

Declarations This work has been partially supported by the national project granted by the Ministry of Science, Innovation and Universities (Spain) with reference RTI2018–099942-B-I00 and by the regional project (ref: SBPLY/17/180501/000495) granted by the regional government (JCCM) and the European Regional Development Funds (FEDER).

Conflicts of interest/Competing interests On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Alepis E, Virvou M (2012) Multimodal object oriented user interfaces in mobile affective interaction. Multimed Tools Appl 59(1):41–63
2. Arroyo I, Cooper DG, Burleson W, Woolf BP, Muldner K, Christopherson R (2009) Emotion sensors go to school. Front Artificial Intel App 200(1):17–24. https://doi.org/10.3233/978-1-60750-028-5-17

3. Blackwell AF and Green TRG (2000) “A Cognitive Dimensions Questionnaire Optimised for Users,” Proc. 12th Work. Psychol. Program. Interes. Gr., no. April, pp. 137–154.
4. Calvo RA, D’Mello S (2010) Affect detection: an interdisciplinary review of models, methods, and their applications. IEEE Trans Affect Comput 1(1):18–37
5. Cambria E, Grassi M, Hussain A, Havasi C (2012) Sentic Computing for social media marketing. Multimed Tools Appl 59(2):557–577
6. Chao X, Zhiyong F (2008) A trusted affective model approach to proactive health monitoring system. Proc - 2008 Intern Sem Fut BioMed Inform Engin, FBIE 2008:429–432. https://doi.org/10.1109/FBIE.2008.52
7. Chen J, Hu B, Li N, Mao C, and Moore P (2013) “A multimodal emotion-focused e-health monitoring support system,” in Proceedings - 2013 7th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2013, pp. 505–510, https://doi.org/10.1109/CISIS.2013.92.
8. Chen LS, Huang TS, Miyasato T, and Nakatsu R 1998 “Multimodal human emotion/expression recognition,” in Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 366–371.
9. Clarke S (2020) “Measuring API Usability”, Dr. Dobb's: The World of Software Development, May 01, 2004. Accessed on: February 12, Available at:https://www.drdobbs.com/windows/measuring-api-usability/184405654
10. Clarke S, Becker C (2003) Using the Cognitive Dimensions Framework to evaluate the usability of a class library. Proc First Jt Conf EASE PPIG, no. April:359–366
11. Dai W, Liu Z, Yu T, and Fung P (2020) “Modality-transferable emotionembeddings for low-resource multimodal emotion recognition,”.
12. Darekar RV, Dhande AP (2016) Enhancing effectiveness of emotion detection by multimodal fusion of speech parameters. Intern Conf Electri, Electron Optimi Tech, ICEEOT 2016:3242–3246. https://doi.org/10.1109/ICEEOT.2016.7755303
13. S. D’Mello, A. Graesser, and R. W. Picard, “Toward an affect-sensitive auHERAutor,” IEEE Intell Syst, vol. 22, no. 4, pp. 53–61, Jul. 2007, https://doi.org/10.1109/MIS.2007.79.
14. D’Mello SK, Kory J(2015) “A review and meta-analysis of multimodal affect detection systems,” ACM Computing Surveys, vol. 47, no. 3. Association for Computing Machinery, 01-Feb-2015.
15. Ekman P (1999) Basic emotions. In: Handbook of cognition and emotion, vol ch. 3. John Wiley & Sons, New York, pp 45–60
16. Express, “Fast, unopinionated, minimalist web framework for Node.js”(2020). Accessed on: April 10th, 2020. Available at: https://expressjs.com/
17. Fabien Mäel (2019) “Multimodal-Emotion-Recognition”, June 28, 2019. Accessed on: March 31, 2020. Available: https://github.com/maelfabien/Multimodal-Emotion-Recognition
18. Garcia-Garcia, Jose Maria, “HERA system: Three-level multimodal emotion recognition framework to detect emotions combining different inputs with different formats. Accessed on: April 10th 2020. Available at: https://github.com/josemariagarcia95/hera-system
19. Garcia-Garcia JM, Penichet VMR, and Lozano MD (2017) “Emotion detection: a technology review,” in Proceedings of the XVIII International Conference on Human Computer Interaction - Interacción ‘17, pp. 1–8.
20. Garcia-Garcia JM, Penichet VMR, Lozano MD, Garrido JE, Lai-Chong Law E (2018) Multimodal affective computing to enhance the user experience of educational software applications. Mob Inf Syst 2018(10):10. https://doi.org/10.1155/2018/8751426
21. Garcia-Garcia JM, Cabañero M e del M, Penichet VMR, and Lozano MD(2019) “EmoTEA: Teaching Children with Autism Spectrum Disorder to Identify and Express Emotions,” in Proceedings of the XX International Conference on Human Computer Interaction - Interacción ‘19, pp. 1–8, https://doi.org/10.1145/3335595.3335639.
22. Gilleade KM, Alan D, and Allanson J (1997) “Affective videogames and modes of affective gaming: assist me, challenge me, emote me,” 2005, .D. L. Hall and J. Llinas, “An introduction to multisensor data fusion,” Proc IEEE, vol. 85, no. 1, pp. 6–23.
23. Gonzalez-Sanchez J, Chavez-Echeagaray M-E, Atkinson R, Burleson W (2011) Affective computing meets design patterns: A pattern-based model for a multimodal emotion recognition framework. Proc 16th Eur Conf Pattern Lang Programs - Eur 11, no. July:1–11. https://doi.org/10.1145/2396716.2396730
24. J. Gonzalez-Sanchez, M. E. Chavez-Echeagaray, R. Atkinson, and W. Burleson, “ABE: An agent-based software architecture for a multimodal emotion recognition framework,” Proc - 9th Work IEEE/IFIP Conf Softw Archit WICSA 2011, no. May 2014, pp. 187–193, 2011, https://doi.org/10.1109/WICSA.2011.32
25. Green TRG (1989) Cognitive dimensions of notations. In: Sutcliffe A, Macaulay L (eds) People and computers V. Cambridge University Press, Cambridge, UK, pp 443–460
26. Green TRG, Petre M (1996) Usability analysis of visual programming environments: a ‘cognitive dimensions’ framework. J Vis Lang Comput 7(2):131–174. https://doi.org/10.1006/jvlc.1996.0009

27. Gupta SK, Ashwin TS, Guddeti RMR (2019) Students' affective content analysis in smart classroom environment using deep learning techniques. Multimed Tools Appl 78(18) Multimedia Tools and Applications:25321–25348
28. A. G. Hauptmann and P. McAvinney, "Gestures with speech for graphic manipulation," Int J Man Mach Stud, vol. 38, no. 2, pp. 231–249, Feb. 1993.
29. Hung JC-S, Chiang K-H, Huang Y-H, Lin K-C (2017) Augmenting teacher-student interaction in digital learning through affective computing. Multimed Tools Appl 76(18) Multimedia Tools and Applications:18361–18386
30. Jaiswal S, Virmani S, Sethi V, De K, Roy PP (2019) An intelligent recommendation system using gaze and emotion detection. Multimed Tools Appl 78(11):14231–14250
31. Jaques N, Conati C, Harley JM, Azevedo R (2014) "Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System," in Intelligent Tutoring Systems. Springer, Cham, pp 29–38
32. Jarraya SK, Masmoudi M, Hammami M (2021) A comparative study of autistic children emotion recognition based on Spatio-temporal and deep analysis of facial expressions features during a meltdown crisis. Multimed Tools Appl 80(1):83–125
33. Khanh TLB, Kim S-H, Lee G, Yang H-J, Baek E-T (2021) Korean video dataset for emotion recognition in the wild. Multimed Tools Appl 80(6):9479–9492
34. Kleinginna PRJ, Kleinginna AM (1981) A categorized list of emotion definitions, with suggestions for a consensual definition. Motiv Emot 5(3):263–291
35. Kołakowska A, Landowska A, Szwoch M, Szwoch W, Wróbel M (2015) Modeling emotions for affect-aware applications. In: Wrzycza S (ed) Information systems development and applications. Faculty of Management University of Gdańsk, Poland, pp 55–69
36. Koelstra S, Muhl C, Soleymani M, Jong-Seok Lee A, Yazdani T, Ebrahimi T, Pun A, Nijholt IP (2012) DEAP: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31
37. Kossaifi, Jean, Robert Walecki, Yannis Panagakis, Jie Shen, Maximilian Schmitt, Fabien Ringeval, Jing Han et al (2019) "SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild." IEEE Transactions on Pattern Analysis and Machine Intelligence.
38. Kumar A, Garg G (2019) Sentiment analysis of multimodal twitter data. Multimed Tools Appl 78(17): 24103–24119
39. Lewis JR (2018) Measuring perceived usability: the CSUQ, SUS, and UMUX. Int J Hum Comput Interact 34(12):1148–1156. https://doi.org/10.1080/10447318.2017.1418805
40. Landowska A (2018) Towards new mappings between emotion representa-tion models. Appl Sci 8(2):274
41. Lewis JR (1995) IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Interact 7:57–78
42. Li Z, Fan Y, Jiang B, Lei T, Liu W (2019) A survey on sentiment analysis and opinion mining for social multimedia. Multimed Tools Appl 78(6):6939–6967
43. Mansoorizadeh M, Moghaddam Charkari N (2010) Multimodal information fusion application to human emotion recognition from face and speech. Multimed Tools Appl 49(2):277–297
44. Maat L, Pantic M (2007) Gaze-X: Adaptive, affective, multimodal interface for single-user office scenarios. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 4451 LNAI: 251–271. https://doi.org/10.1007/978-3-540-72348-6_13
45. Martin B, Hanington B (2012) Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Publishers, Beberly (Massachusetts), pp 204–205
46. Mehrabian A, Russell JA (1974) An approach to environmental psychology. The MIT press
47. Mittal T, Guhan P, Bhattacharya U, Chandra R, Bera A, Manocha D (2020, 2020) EmotiCon: Context-Aware Multimodal Emotion Recognition Using Frege's Principle. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), Seattle, WA, USA, pp 14222–14231. https://doi.org/10.1109/CVPR42600.2020.01424
48. Nielsen J, Landauer T (1993) A mathematical model of the finding of usability problems. Proceedings of the Interact'93 and CHI'93 Conference on Human Factors in Computing systems; 1993 Apr. ACM, Amsterdam, the Netherlands. New York, pp 24–29
49. Osman H. Al and Falk TH (2017) "Multimodal Affect Recognition: Current Approaches and Challenges," in Emotion and Attention Recognition Based on Biological Signals and Images, InTech.
50. Oviatt S, DeAngeli A, and Kuhn K (1997) "Integration and synchronization of input modes during multimodal human-computer interaction," in Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97, pp. 415–422.
51. Oehl M, Siebert FW, Tews T-K, Höger R, Pfister H-R (2011) Improving human-machine interaction - A non-invasive approach to detect emotions in car drivers. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 6763 LNCS, no. PART 3:577–585
52. Pantic M, Sebe N, Cohn JF, Huang T (2005) Affective multimodal human-computer interaction. Proc 13th ACM Int Conf Multimedia, MM 2005 , no. January:669–676

53. Patwardhan AS (2018) “Multimodal mixed emotion detection,” in Proceedings of the 2nd International Conference on Communication and Electronics Systems, ICCES 2017, 2018, vol., pp. 139–143, https://doi.org/10.1109/CESYS.2017.8321250.
54. Picard RW (1995) Affective Computing. MIT Press 321:1–16
55. Poria S, Cambria E, Bajpai R, and Hussain A (2017) “A review of affective computing: from unimodal analysis to multimodal fusion,” Inf. Fusion.
56. Pyeon Myeongjang (2018) “IEMo: web-based interactive multimodal emotion recognition framework”, Abril 30, 2018. Accessed on: April 28, 2020. Available at: https://github.com/mjpyeon/IEMo
57. Ringeval F, Eyben F, Kroupi E, Yuce A, Thiran JP, Ebrahimi T, Lalanne D, Schuller B (2015) Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. Pattern Recogn Lett 66:22–30
58. Rousidis D, Koukaras P, Tjortjis C (2020) Social media prediction: a literature review. Multimed Tools Appl 79(9–10):6279–6311
59. Sekhavat YA, Sisi MJ, and Roohi S (2020) “Affective interaction: using emotions as a user interface in games”, Multimedia Tools and Applications. Multimedia Tools and Applications, Affective interaction: Using emotions as a user interface in games.
60. Sethu V, Provost EM, Epps J, Busso C, Cummins N, and Narayanan S 2019 “The ambiguous world of emotion representation,”.
61. Silva L. C. De, Miyasato T, and Nakatsu R (1997) “Facial emotion recognition using multi-modal information,” Proc. ICICS, 1997 Int. Conf. Information, Commun. Signal Process. Theme Trends Inf. Syst. Eng. Wirel. Multimed. Commun. (Cat. No.97TH8237), vol. 1, no. May 2014, pp. 397–401.
62. L. C. De Silva, Pei Chi Ng (2000) “Bimodal emotion recognition,” in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), pp. 332–335.
63. Siriwardhana S, Kaluarachchi T, Billinghurst M, Nanayakkara S (2020) Multimodal emotion recognition with transformer-based self supervised feature fusion. IEEE Access 8:176274–176285. https://doi.org/10.1109/ACCESS.2020.3026823
64. W3C, emotion markup language, (May 22, 2014). Accessed on: February 17th, 2020. Available: https://www.w3.org/TR/emotionml/
65. W3C, multimodal interaction framework, multimodal interaction working group, (May 06, 2003). Accessed on: February 17th, 2020. Arvailable: https://www.w3.org/TR/mmi-framework/
66. Wang Z, Ho S-B, Cambria E (2020) A review of emotion sensing: categorization models and algorithms. Multimed Tools Appl 79(47–48):35553–35582
67. Wijayarathna C, Arachchilage NAG, Slay J (2017) “Using Cognitive Dimensions Questionnaire to Evaluate the Usability of Security APIs,” no. 2004.
68. Woolf B, Woolf B, Burelson W, Arroyo I(2007) “Emotional Intelligence for Computer Tutors,” Suppl. Proc. 13TH Int. Conf. Artif. IN-TELLIGENCE Educ. (AIED 2007), (PP, pp. 6–15.
69. Yamauchi T (2013) “Mouse Trajectories and State Anxiety: Feature Selection with Random Forest,” in 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, pp. 399–404.
70. Zhao S et al. (2020) "Discrete Probability Distribution Prediction of Image Emotions with Shared Sparse Learning," in IEEE Transactions on Affective Computing, vol. 11, no. 4, pp. 574–587, 1 Oct.-Dec, https://doi.org/10.1109/TAFFC.2018.2818685.
71. Zhao S, Gholaminejad A, Ding G, Gao Y, Han J, Keutzer K (2019) Personalized emotion recognition by personality-aware high-order learning of physiological signals. ACM Trans Multimed Comput Commun Appl 15, 1s, article 14, (February 2019):18. https://doi.org/10.1145/3233184
72. Zhao S, Ding G, Gao Y, Han J(2017) “Approximating discrete probability distribution of image emotions by multi-modal features fusion,“in Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 4669–4675
73. Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J. F. Cohn, Q. Ji, and L. Yin, “Multimodal spontaneous emotion Corpus for human behavior analysis”, 2016.
74. Zhang S, Wu Z, Meng HM, Cai L (2010) Facial expression synthesis based on emotion dimensions for affective talking avatar. Smart Innov Syst Technol 2010(1):109–132. https://doi.org/10.1007/978-3-642-12604-8_6
75. W. L. Zheng, W. Liu, Y. Lu, B. L. Lu, and A. Cichocki, “EmotionMeter: a multimodal framework for recognizing human emotions,” IEEE Trans Cybern, vol. 49, no. 3, pp. 1110–1122, Mar. 2019, https://doi.org/10.1109/TCYB.2018.2797176.

Affiliations

Jose Maria Garcia-Garcia[1] • **Maria Dolores Lozano**[2] • **Victor M. R. Penichet**[2] • **Effie Lai-Chong Law**[3]

Maria Dolores Lozano
Maria.Lozano@uclm.es

Victor M. R. Penichet
Victor.Penichet@uclm.es

Effie Lai-Chong Law
lai-chong.law@durham.ac.uk

[1] Informatics Research Institute of Albacete, 02006 Albacete, Spain

[2] Department of Computing Systems, Universidad de Castilla-La Mancha, 02006 Albacete, Spain

[3] Department of Computer Science, Durham University, Durham DH1 3LE, UK

Chapter 4

Conclusions and future work

This chapter presents the main conclusions that may be drawn from the research performed. In addition, some future research lines that could be explored starting from the findings of this Thesis are described at the end of the chapter.

4.1 Conclusions

This PhD Thesis presented as compendium of publications has contributed with new interaction mechanisms which integrate affective technology and Human-Computer Interaction techniques, which have been applied and assessed in different application domains. The final results consist of a series of prototypes which have been enriched with emotion detection capabilities, granting them the ability to modify their behaviour automatically according to the users' mood, to log affective information from the users for later processing, etc. With these proposals, we can consider the main goal of this thesis, as well as the specific goals, fulfilled.

It is important to highlight that, even though this thesis by compendium is endorsed by three publications in international scientific journals ([16][19][20]), some others works outside of these previous papers have been carried out as well, such us conferences publications [16][18], Bachelor theses [13][44] and publications which have not been published yet.

The contributions which endorse this thesis by compendium can be organised in a set of steps:

1. A literature review has been carried out to gain insight of the different elements, entities and disciplines which participate in Affective Computing. This review allowed us to understand the current trends existent in the field and to discover potential gaps in the research being carried out. This knowledge allowed us to build a strong bedrock from where we would building our next projects.

2. While studying the big picture of *AC*, we decided to pay attention to the type of publication which was attracting the most attention in the field. After completing the aforementioned

review, we discovered that most of the real-world applications being built belong to the fields of e-learning, e-health and marketing. This informations allowed us to consider a multiple view perspective when building our own proposals.

3. Using the information gathered during the previous review, we developed several prototypes integrating mechanisms from *AC* and *HCI* techniques, offering new interaction mechanisms through the use of emotion detection.

4. After building the aforementioned prototypes, we decided to take a step forward in the proposal of affective prototypes using this new acquired experience. Since part of designing and building affective software involves integrating emotion detection into a bigger system, we decided to propose an architecture to assist in this task. This architecture, which also supports the creation of multimodal emotion detectors, has the potential to support researchers and developers in their tasks to add emotion detection capabilities to a system.

5. The previous works led us to ponder about the impact that affective mechanisms could have on software quality, and how this could be measured. While existent quality models are designed with high levels of abstractions, so they can be adapted to virtually any software, they fail to capture some of the idiosyncrasies of affective technologies. In order to overcome this obstacle, we proposed en extension of the quality model defined in the ISO 25010, to support developers and researchers in the quality assessment of affective applications.

4.2 Future work

We have identified different ways in which the research outcomes presented in this Thesis could be extended. Firstly, the most urgent task is to finish the paper about our quality model extension proposal. This paper, which has been developed with the collaboration of Professor Per Ola Kristensson from the University of Cambridge, is also the result of a research stay which took place from April 2022 to July 2022 in Cambridge (United Kingdom), at the Intelligent Interactive Systems research group.

Secondly, we are studying the possibility of designing a framework to assist researchers in the development of affective prototypes oriented to e-learning. This proposal, which has not been completed refined yet, should include support to create users, add emotion detection service, connect to other devices over Bluetooth, HTTP requests and other communication protocols, etc.

Finally, one of our works which has not been published yet is a proposal to include emotion detection in a system built to perform remote physical therapy using Kinect v2. This proposal exploits the relationship between emotion, motivation and recovery speed, and it is still being developed right now.

Capítulo 5

Conclusiones y trabajo futuro

Este capítulo presenta las conclusiones principales que se extraen de los trabajos realizados. Al final de este capítulo, se comentan también algunas de las líneas de investigación futuras que se construirán sobre los hallazgos y conclusiones extraídos de esta tesis doctoral.

5.1 Conclusiones

En esta tesis por compendio se han presentado diversas soluciones y propuestas para constituir nuevos mecanismos de interacción que han sido enriquecidos con técnicas de Computacion Afectiva, expandiendo las capacidades de aplicaciones y sistemas informáticos para dotarlos de capacidades afectivas que permitan llevar a cabo nuevos tipos de interacción y ofrecer nuevas experiencias de usuario. Así, se han cumplido tanto el objetivo principal de esta tesis como sus objetivos específicos (ver Sección 1.3), produciéndose una serie de resultados que satisfacen dichos objetivos (ver Capítulo 2).

Es importante destacar que este compendio de publicaciones está avalado por tres publicaciones ([16][19][20]), sin perjuicio de otras publicaciones [16][18], Trabajos de Fin de Grado [13][44] y artículos en proceso de publicación que también se han producido en el marco de esta tesis.

De esta forma, las aportaciones que avalan esta tesis pueden definirse en esta serie de pasos:

1. Se ha realizado una revisión de la literatura que ha permitido conocer el panorama actual de la Computación Afectiva y entender las tendencias de investigación actuales en congresos y revistas sobre el tema, así como los principales interesados y promotores de esta nueva disciplina de la informática. Este paso nos permite establecer unas bases sólidas sobre las que construir las aportaciones posteriores.

2. Una vez que se acotó el campo, se realizó un estudio sobre las distintas aplicaciones que más atención estaban atrayendo en el sector, que resultaron ser, entre otras, la educación y la sanidad, seguidas de cerca por el marketing y el diseño. Analizar las distintas formas de aplicar

la *CA* en distintos campos dio una visión de perspectiva múltiple que fui muy útil a la hora de diseñar las distintas aportaciones de esta tesis.

3. Usando el conocimiento adquirido con los dos hitos anteriores, se desarrollaron varios sistemas interactivos que integraban mecanismos propios de la *CA* con técnicas de *IPO*, lo que dio lugar a un abanico de propuestas que integraban distintas formas de detección de emociones para ofrecer nuevos mecanismos de interacción.

4. Tras el desarrollo de diversos prototipos afectivos, se propuso una arquitectura para la implementación de detectores de emociones multimodales, esto es, detectores detectan emociones en más de un canal a la vez. Dada la dificultad que la implementación de estos detectores suele conllevar, en esta tesis se propuso e implementó una arquitectura de detector multimodal para aliviar esta dificultad.

5. El desarrollo de los hitos anteriores supuso reflexionar sobre la forma de evaluar la calidad de aplicaciones afectivas, y cómo los modelos existentes fallan en capturar algunos aspectos importantes de estas aplicaciones, sin perjuicio de que sean totalmente válidos para medir la calidad de software afectivo. Sin embargo, en pos de ofrecer un modelo de calidad que no borre características importantes de posibles sistemas afectivos, se realizó una extensión del modelo de calidad propuesto en la ISO 25010 para dar cabida a la evaluación de sistemas que integren *CA*.

5.2 Trabajo futuro

En lo que respecta a trabajo futuro, la tarea más inmediata consiste en finalizar la publicación sobre la extensión del modelo de calidad de la ISO 25010. Esta publicación constituye además el resultado de haber realizado una estancia de tres meses entre abril y julio de 2022 con el Dr. Per Ola Kristensson, en la Universidad de Cambridge, en el grupo de investigación Intelligent Interactive Systems.

En segundo lugar, se ha planteado desarrollar un marco de trabajo para el desarrollo de aplicaciones afectivas enfocadas al sector de la educación, ofreciendo funcionalidad básica como la gestión de usuarios, el acceso a sensores y detectores a través de distintos protocolos de comunicación, etc., agilizando el proceso de desarrollo de herramientas afectivas en el campo de la educación.

Por último, uno de los trabajos pendientes de publicación consiste en una ampliación de una aplicación para telerrehabilitación física para dotarla de la capacidad de leer las emociones de los pacientes durante sus rutinas de ejercicios y analizar el progreso de estos en función de las emociones experimentadas a lo largo del proceso de rehabilitación.

Bibliografía

[1] Babak Joze Abbaschian, Daniel Sierra-Sosa, and Adel Elmaghraby. Deep learning techniques for speech emotion recognition, from databases to models. *Sensors*, 21(4):1249, 2021.

[2] Siddeeq Y.; Sadeeq Mohammed A. M. Abdullah, Sharmeen M Saleem; Ameen and Subhi Zeebaree. Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*, 2(02):52–58, 2021.

[3] Sharifa Alghowinem. From joyous to clinically depressed: Mood detection using multimodal analysis of a person's appearance and speech. pages 648–654, 09 2013. doi: 10.1109/ACII.2013.113.

[4] Renan Vinicius Aranha, Cleber Gimenez Correa, and Fatima L.S. Nunes. Adapting software with affective computing: A systematic review. *IEEE Transactions on Affective Computing*, 12: 883–899, 2021. ISSN 19493045. doi: 10.1109/TAFFC.2019.2902379.

[5] Katharine Howard Blocher. Affective social quest (asq): Teaching emotion recognition with interactive media and wireless expressive toys. In *Master's Thesis for Master of Science in Media Technology Massachusetts Institute of Technology, MIT*. MIT, 1999.

[6] Rafael A. Calvo and Sidney D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1:18–37, 2010. ISSN 19493045. doi: 10.1109/T-AFFC.2010.1.

[7] Rodrigo S Cañibano, Santino Castagno, Mariano Conchillo, Guillermo Chiarotto, Claudia Rozas, Claudio Zanellato, Cristina Orlandi, and Javier Balladini. Towards a resilient e-health system for monitoring and early detection of severity in hospitalized patients during a pandemic. In *X Jornadas de Cloud Computing, Big Data & Emerging Topics (La Plata, 2022)*, 2022.

[8] Delphine Caruelle, Poja Shams, Anders Gustafsson, and Line Lervik-Olsen. Affective computing in marketing: Practical implications and research opportunities afforded by emotionally intelligent machines. *Marketing Letters*, 33(1):163–169, 2022. ISSN 0923-0645. doi: 10.1007/s11002-021-09609-0.

[9] Jing Chen, Bin Hu, Na Li, Chengsheng Mao, and Philip Moore. A multimodal emotion-focused e-health monitoring support system. pages 505–510, 2013. ISBN 9780769549927. doi: 10.1109/CISIS.2013.92.

[10] Joanne M. Daly, Britton W. Brewer, Judy L. Van Raalte, Albert J. Petitpas, and Joseph H. Sklar. Cognitive appraisal, emotional adjustment, and adherence to rehabilitation following knee surgery. *Journal of Sport Rehabilitation*, 4(1):23 – 30, 1995. doi: 10.1123/jsr.4.1.23. URL https://journals.humankinetics.com/view/journals/jsr/4/1/article-p23.xml.

[11] Nidal Daou, Ryma T Hady, and Claire L Poulson. Teaching children with autism spectrum disorder to recognize and express emotion: A review of the literature. *International Electronic Journal of Elementary Education*, 9(2):419–432, 2016.

[12] Geraldine Dawson and Kathleen Zanolli. Early intervention and brain plasticity in autism. *Autism: Neural bases and treatment possibilities*, 251:266–80, 2003.

[13] Luis del Moral Tébar. Mejora por medio de detección de emociones y visualización de estadísticas de una herramienta de rehabilitación basada en movimiento. Bachelor's thesis, Universidad de Castilla-La Mancha, 2021.

[14] Michalis Feidakis, Maria Rangoussi, Panagiotis Kasnesis, Charalampos Patrikakis, Dimitrios Kogias, and Angelos Charitopoulos. Affective assessment in distance learning: A semi-explicit approach. *International Journal of Technologies in Learning*, 26:19–34, 04 2019. doi: 10.18848/2327-0144/CGP/v26i01/19-34.

[15] Jose Maria Garcia-Garcia, Victor M R Penichet, and Maria D Lozano. Emotion detection: a technology review. pages 1–8. ACM Press, 2017. ISBN 9781450352291. doi: 10.1145/3123818.3123852. URL http://dl.acm.org/citation.cfm?doid=3123818.3123852.

[16] Jose Maria Garcia-Garcia, Víctor M R Penichet, María Dolores Lozano, Juan Enrique Garrido, and Effie Lai-Chong Law. Multimodal affective computing to enhance the user experience of educational software applications. *Mobile Information Systems*, 2018:10, 2018. doi: 10.1155/2018/8751426. URL https://doi.org/10.1155/2018/8751426.

[17] Jose Maria Garcia-Garcia, María del Mar Cabañero, Victor M. R. Penichet, and María D. Lozano. Emotea: Teaching children with autism spectrum disorder to identify and express emotions. pages 1–8. ACM Press, 2019. ISBN 9781450371766. doi: 10.1145/3335595.3335639. URL http://dl.acm.org/citation.cfm?doid=3335595.3335639.

[18] Jose Maria Garcia-Garcia, María del Mar Cabañero, Victor M. R. Penichet, and María D. Lozano. Emotea: Teaching children with autism spectrum disorder to identify and express emotions. pages 1–8. ACM Press, 2019. ISBN 9781450371766. doi: 10.1145/3335595.3335639. URL http://dl.acm.org/citation.cfm?doid=3335595.3335639.

[19] Jose Maria Garcia-Garcia, Víctor M R Penichet, María Dolores Lozano, and Anil Fernando. Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions. *Universal Access in the Information Society*, page 809–825, 2021. ISSN 1615-5289. doi: 10.1007/s10209-021-00818-y. URL https://doi.org/10.1007/s10209-021-00818-y.

[20] Jose Maria Garcia-Garcia, Maria Dolores Lozano, Victor M. R. Penichet, and Effie Lai-Chong Law. Building a three-level multimodal emotion recognition framework. *Multimedia Tools and Applications*, 6 2022. ISSN 1380-7501. doi: 10.1007/s11042-022-13254-8.

[21] D.L. Hall and J. Llinas. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85:6–23, 1997. ISSN 00189219. doi: 10.1109/5.554205. URL http://ieeexplore.ieee.org/document/554205/.

[22] Tina Hascher. Learning and emotion: Perspectives for theory and research. *European Educational Research Journal*, 9(1):13–28, 2010. doi: 10.2304/eerj.2010.9.1.13. URL https://doi.org/10.2304/eerj.2010.9.1.13.

[23] Uwe Herwig, Tina Kaffenberger, Lutz Jäncke, and Annette B. Brühl. Self-related awareness and emotion regulation. *NeuroImage*, 50(2):734–741, 2010. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2009.12.089. URL https://www.sciencedirect.com/science/article/pii/S1053811909013780.

[24] John Wesley Hill. *Deep Learning for Emotion Recognition in Cartoons*. PhD thesis, Lincoln School of Computer Science, University of Lincoln, 2017. URL https://hako.github.io/dissertation/.

[25] ISO 25010:2010(E). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. Standard, International Organization for Standardization, March 2011.

[26] Vaishali M Joshi and Rajesh B Ghongade. Idea: Intellect database for emotion analysis using eeg signal. *Journal of King Saud University-Computer and Information Sciences*, 2020.

[27] Dana Kirsch. *The Affective Tigger : a study on the construction of an emotionally reactive toy*. PhD thesis, 1999.

[28] Shashidhar G Koolagudi, Sudhamay Maity, Vuppala Anil Kumar, Saswat Chakrabarti, and K Sreenivasa Rao. Iitkgp-sesc: speech database for emotion analysis. In *International conference on contemporary computing*, pages 485–492. Springer, 2009.

[29] Andrej Luneski, Panagiotis D Bamidis, and Madga Hitoglou-Antoniadou. Affective computing and medical informatics: state of the art in emotion-aware medical applications. *Studies in health technology and informatics*, 136:517, 2008.

[30] Wafa Mellouk and Wahida Handouzi. Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science*, 175:689–694, 2020.

[31] Daniel S. Messinger, Leticia Lobo Duvivier, Zachary E. Warren, Mohammad Mahoor, Jason Baker, Anne S. Warlaumont, and Paul Ruvolo. Affective computing, emotional development, and autism. 1 2015. doi: 10.1093/OXFORDHB/9780199942237.013.012.

[32] Michael Oehl, Felix W Siebert, Tessa-Karina Tews, Rainer Höger, and Hans-Rüdiger Pfister. Improving human-machine interaction–a non invasive approach to detect emotions in car drivers. In *International conference on human-computer interaction*, pages 577–585. Springer, 2011.

[33] Hussein Al Osman and Tiago H. Falk. Multimodal affect recognition: Current approaches and challenges. *Emotion and Attention Recognition Based on Biological Signals and Images*, 2 2017. doi: 10.5772/65683. URL http://www.intechopen.com/books/emotion-and-attention-recognition-based-on-biological-signals-and-images/multimodal-affect-recognition-current-approaches-and-challenges.

[34] Elizabeth A. Phelps. Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, 57(1):27–53, 2006. ISSN 0066-4308. doi: 10.1146/annurev.psych.56.091103.070234.

[35] Rosalind W. Picard. Affective computing. *MIT press*, pages 1–16, 1995. ISSN 09269630. doi: 10.1007/BF01238028. URL papers3://publication/uuid/9C02FCAE-FE2E-4D2C-9707-766804777DC9.

[36] Filip Radulovic, Raúl García-Castro, and Asunción Gómez-Pérez. Semquare - an extension of the square quality model for the evaluation of semantic technologies. *Computer Standards and Interfaces*, 38:101–112, 2015. ISSN 09205489. doi: 10.1016/j.csi.2014.09.001.

[37] Jocelyn Riseberg, Jonathan Klein, Raul Fernandez, and Rosalind W. Picard. Frustrating the user on purpose: using biosignals in a pilot study to detect the user's emotional state. *CHI 98 Conference*, pages 1–2, 1998. doi: 10.1145/286498.286715. URL http://dl.acm.org/citation.cfm?id=286715.

[38] Jesús Joel Rivas, Felipe Orihuela-Espina, Lorena Palafox, Nadia Bianchi-Berthouze, María del Carmen Lara, Jorge Hernández-Franco, and Luis Enrique Sucar. Unobtrusive inference of affective states in virtual rehabilitation from upper limb motions: A feasibility study. *IEEE Transactions on Affective Computing*, 11(3):470–481, 2020. doi: 10.1109/TAFFC.2018.2808295.

[39] O C Santos. Emotions and personality in adaptive e-learning systems: an affective computing perspective. *Emotions and Personality in Personalized Services*, pages 263–285, 2016. doi: 10.1007/978-3-319-31413-6_13. URL http://link.springer.com/chapter/10.1007/978-3-319-31413-6_13.

[40] Klaus R. Scherer. Expression of emotion in voice and music. *Abstract Journal of Voice*, 9: 235–248, 1995.

[41] L.C. De Silva and Pei Chi Ng. Bimodal emotion recognition. pages 332–335. IEEE Comput. Soc, 2000. ISBN 0-7695-0580-5. doi: 10.1109/AFGR.2000.840655. URL http://ieeexplore.ieee.org/document/840655/.

[42] L.C. De Silva, T. Miyasato, and R. Nakatsu. Facial emotion recognition using multi-modal information. *Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat. No.97TH8237)*, 1:397–401, 1997. doi: 10.1109/ICICS.1997.647126. URL http://ieeexplore.ieee.org/document/647126/.

[43] Nidhi Sinha. Affective computing and emotion-sensing technology for emotion recognition in mood disorders. In *Enhanced Telemedicine and e-Health*, pages 337–360. Springer, 2021.

[44] Alejandro José Martínez Sánchez. Detección y visualización de emociones en alumnos en un entorno educativo. Bachelor's thesis, Universidad de Castilla-La Mancha, 2020.

[45] Yan Wang, Wei Song, Wei Tao, Antonio Liotta, Dawei Yang, Xinlei Li, Shuyong Gao, Yixuan Sun, Weifeng Ge, Wei Zhang, and et al. A systematic review on affective computing: emotion models, databases, and recent advances. *Information Fusion*, 83-84:19–52, 2022. ISSN 1566-2535. doi: 10.1016/j.inffus.2022.03.009.

[46] Elaheh Yadegaridehkordi, Nurul Fazmidar Binti Mohd Noor, Mohamad Nizam Bin Ayub, Hannyzzura Binti Affal, and Nornazlita Binti Hussin. Affective computing in education: A systematic review and future research. *Computers Education*, 142:103649, 2019. ISSN 0360-1315. doi: https://doi.org/10.1016/j.compedu.2019.103649. URL https://www.sciencedirect.com/science/article/pii/S0360131519302027.

[47] Takashi Yamauchi. Mouse trajectories and state anxiety: feature selection with random forest. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 399–404. IEEE, 2013.

[48] Hanan Makki Zakari, Minhua Ma, and David Simmons. A review of serious games for children with autism spectrum disorders (asd). volume 8778, pages 93–106. Springer Verlag, 2014. ISBN 9783319116228. doi: 10.1007/978-3-319-11623-5_9.