

ANEXO A. GESTIÓN DEL PROYECTO

Este anexo contiene detalles acerca de la gestión del tiempo y el esfuerzo invertido durante el proyecto. En el primer punto se presenta la evolución temporal del proyecto, y en el segundo el esfuerzo y desglose de horas invertidas.

A.1 Gestión del tiempo

Este proyecto ha sido desarrollado desde octubre de 2011 hasta Agosto de 2012, con una dedicación a tiempo parcial durante el primer cuatrimestre y a tiempo completo durante el segundo. En el diagrama de Gantt que aparece en la Fig. A. 1 se puede ver cómo se han distribuido las tareas a lo largo del tiempo.

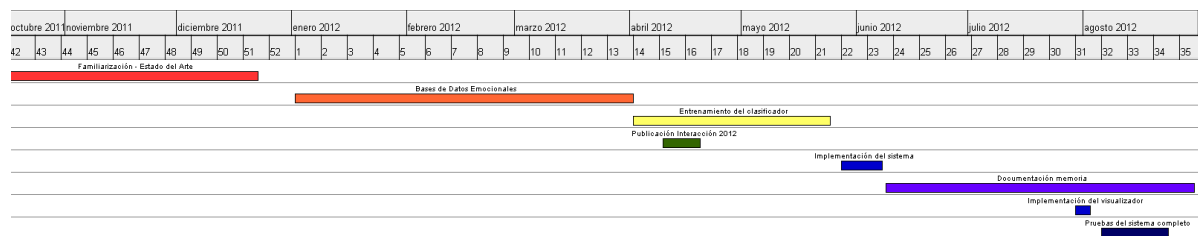


FIG. A. 1. EVOLUCIÓN TEMPORAL DEL PROYECTO

A continuación un pequeño resumen de lo que engloba cada tarea:

Familiarización – Estado del arte [17/10/11 – 23/12/11]. Este proyecto contiene una gran carga de investigación, por lo que ha sido necesaria una inversión de tiempo bastante considerable tanto en la familiarización con el trabajo previo del que se ha partido como en el estudio del estado del arte en la detección automática de emociones. Esta tarea engloba, por tanto, la familiarización y documentación del antiguo sistema, la realización del estado del arte y la prueba y selección de dispositivos de captura

Bases de Datos Emocionales [02/01/12 – 02/04/12]. Una de las principales tareas de este proyecto ha sido la comparación, selección y procesado de vídeos de una Base de Datos Emocional continua. Este proceso ha sido muy costoso dado que muchas partes debían realizarse manualmente y vídeo por vídeo. Esta tarea está compuesta por las sub-tareas de búsqueda de Bases de Datos Emocionales (BDE), selección y evaluación de las más prometedoras, comparación de éstas y finalmente descarga y procesado de la BDE seleccionada.

Entrenamiento del clasificador [02/04/12 – 25/05/12]. La tarea de entrenamiento del clasificador consiste en la extracción de los datos procedentes de los vídeos, la preparación del fichero de clasificación y la realización del *benchmarking* con el software WEKA, donde finalmente se selecciona el clasificador, así como sus entradas y salidas.

Publicación Interacción 2012 [02/04/12 – 20/04/12]. Esta tarea recoge el trabajo realizado para la puesta a punto del artículo que ha sido enviado al XIII congreso internacional de Interacción. Si bien se disponía ya de mucha información por entonces, se trataba del primer artículo de investigación redactado por el autor con lo que fue necesario un esfuerzo considerable.

Implementación del sistema [28/05/12 – 08/06/12]. Una vez entrenado el clasificador, ya se pueden realizar las modificaciones en el sistema previo para migrarlo completamente al continuo. Esta tarea se encarga de las fases de modificación de los programas de extracción de datos y de clasificación, dejando la implementación del visualizador de resultados en otra tarea.

Documentación [09/06/12 – 31/08/2012]. Esta parte se corresponde con el proceso de documentación y redacción de la memoria. Se ha ido trabajando en paralelo con las últimas fases del proyecto.

Implementación del visualizador [30/07/12 – 03/08/12]. El visualizador de resultados corresponde a última fase del proceso de reconocimiento automático de emociones y ha sido renovado completamente respecto al trabajo previo. Esta parte se corresponde al análisis, diseño e implementación del visualizador de resultados.

Pruebas del sistema [6/8/12 – 24/8/12]. Por último, la tarea de pruebas del sistema se encarga de realizar aquellas pruebas que comprueben el correcto funcionamiento del trabajo realizado.

A.2 Esfuerzo invertido

En la realización del proyecto se han invertido un total de **764 horas**. En la Fig. A. 2 se presenta la distribución del tiempo invertido para cada una de las diferentes tareas del proyecto. Se ve claramente que hay tres tareas que son el grueso del proyecto donde ha habido una mayor inversión de horas, correspondientes a la familiarización y estado del arte, a la selección de una BDE y al entrenamiento del clasificador. Una vez realizadas estas tareas más costosas, las demás se han desarrollado más rápidamente.

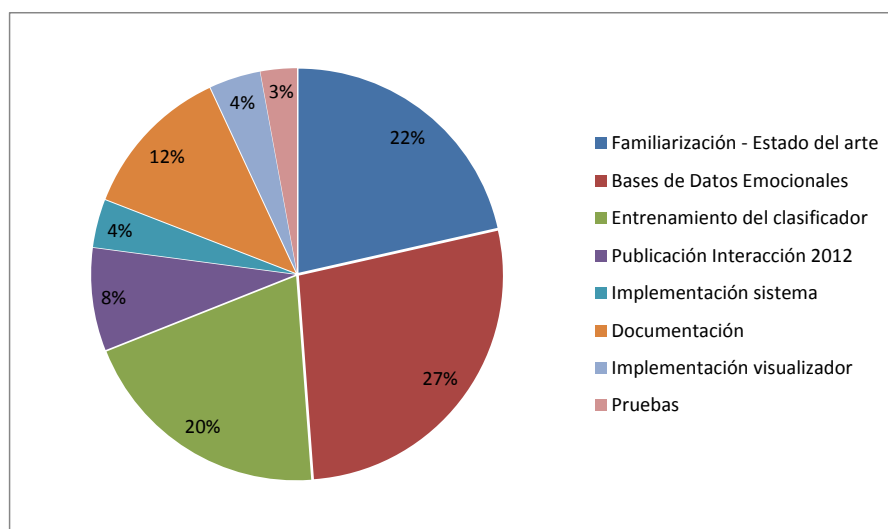


FIG. A. 2. DISTRIBUCIÓN DEL TIEMPO PARA CADA UNA DE LAS DIFERENTES TAREAS DEL PROYECTO

A continuación se muestra en la Tabla A. 1 de una manera más detallada la cantidad de horas invertidas para cada tarea, así como las sub-tareas que la componen.

TABLA A. 1. NÚMERO DE HORAS INVERTIDAS EN LAS TAREAS DEL PROYECTO

Tarea	Nº de horas	
Familiarización - Estado del arte	164	[22%]
Familiarización trabajo previo	29	
Documentación trabajo previo	13	
Estado del arte en reconocimiento de emociones	98	
Pruebas y selección de dispositivos de captura	24	
Bases de Datos Emocionales [BDE]	209	[27%]
Búsqueda de BDE	22	
Selección y evaluación de las BDE más prometedoras	60	
Comparación de BDE	8	
Descarga y procesado de la BDE seleccionada	119	
Entrenamiento del clasificador	154	[20%]
Extracción de datos	118	
Aprendizaje de WEKA	12	
Preparación del fichero de clasificación	6	
Benchmarking	18	
Publicación Interacción 2012	62	[8%]
Implementación del sistema	29	[4%]
Documentación	93	[12%]
Implementación del visualizador	31	[4%]
Pruebas del sistema	22	[3%]
NÚMERO TOTAL DE HORAS	764	

ANEXO B. TECNOLOGÍA Y HERRAMIENTAS UTILIZADAS

Este anexo detalla la tecnología y las herramientas más importantes que han sido utilizadas para el desarrollo del presente Proyecto Fin de Carrera.

B.1 Reconocimiento facial: *FaceAPI*

Existen varias tecnologías de reconocimiento facial. Entre ellas, podemos encontrar Microsoft *Kinect* [MKin], que proporciona un mallado sobre el rostro a tiempo real; o las librerías *OpenCV* [OpCV], que permiten detectar regiones en el rostro. También existen otras tecnologías comerciales de pago como *FaceAPI*, de la empresa *SeeingMachines* [FASM], que proporciona la posición de varios puntos faciales sobre el rostro.

Finalmente, se ha seleccionado *FaceAPI* por ser la tecnología usada en el grupo de investigación y porque ya se contaba con la licencia al comenzar el proyecto. No obstante, la licencia de la que se dispone tiene algunas restricciones como que no reconoce el contorno exacto de los ojos, sino que los estima a partir de los puntos faciales obtenidos de cejas, nariz y labios. Dicha herramienta es un software comercial, cuya licencia viene en forma de memoria USB (también llamado mochila), de modo que no podía emplearse más de un ordenador simultáneamente. *FaceAPI* proporciona unas librerías que ponen a disposición del usuario funciones para el correcto seguimiento facial. Concretamente, la herramienta posibilita:

- Reconocimiento de caras en tiempo real y en 3D.
- Incluye posición y rotación de la cabeza (6 grados de libertad).
- Posición de puntos faciales, incluyendo 16 puntos de control en los labios y 9 en las cejas.

Además, viene con una serie de requerimientos del sistema, que son los siguientes:

- Windows XP SP2 o versión más reciente.
- Intel Core-2 Duo.
- 1GB DRAM
- Se necesitan 10 milisegundos entre medición y medición para el procesamiento de la información.

En la Fig. B. 1 puede observarse una demo del funcionamiento de *FaceAPI*.

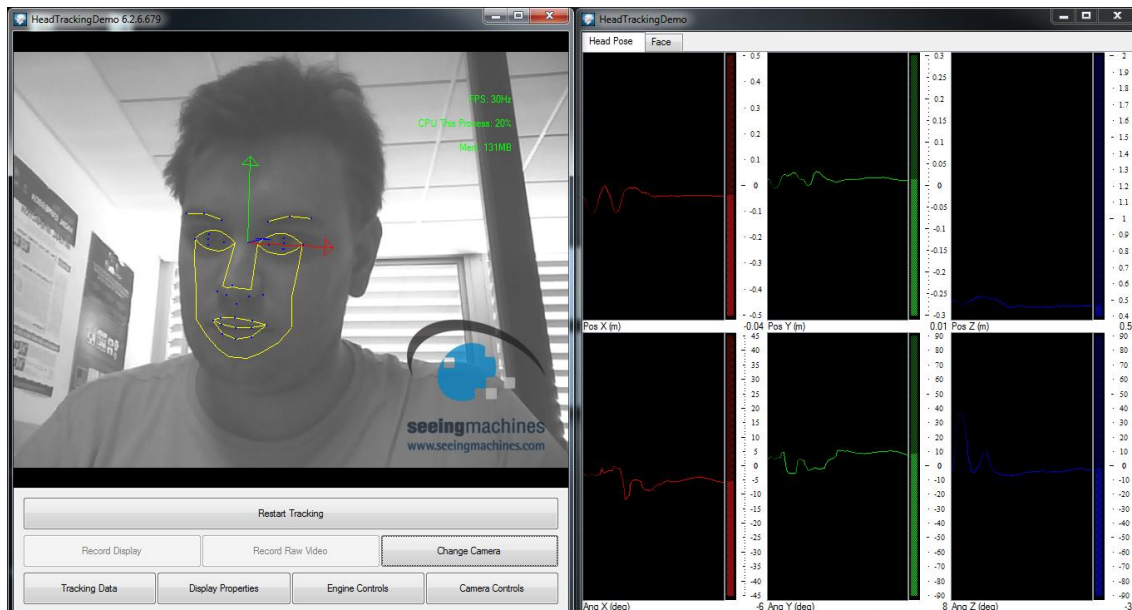


FIG. B. 1. INTERFAZ DE LA DEMO DE RECONOCIMIENTO DEL SOFTWARE *FACEAPI*

B.2 Herramienta de clasificación: *WEKA*

Existen diferentes alternativas a la hora de seleccionar programas de minería de datos y clasificación, como *R[Pro]*, *RapidMiner* [RMin] o *WEKA* [WEKA]. *RapidMiner* se descartó por tener muy pocas prestaciones en su versión gratuita. De entre los dos restantes programas, y tras una comparativa entre ellos [Elop], se seleccionó *WEKA* por tener mayores funcionalidades y, sobre todo, un número mucho mayor de modelos de clasificación.

La herramienta *WEKA* (*Waikato Environment for Knowledge Analysis*) ofrece diferentes funcionalidades como visualización de datos, selección de atributos, herramientas para clasificar y entrenar y realización de validaciones.

La herramienta *WEKA* ha sido usada de dos maneras diferentes: (1) Utilizando la aplicación *WEKA explorer* durante la fase de entrenamiento del clasificador y (2) Utilizando las librerías para el IDE Eclipse que proporciona para Java en la fase de clasificación del sistema de reconocimiento.

B.2.1 *Weka explorer*

En la Fig. B. 2 se muestra el aspecto que ofrece el la aplicación *WEKA explorer*. *WEKA* cuenta con una interfaz gráfica muy potente que permite ver de manera clara y rápida los atributos de la relación del fichero de entrenamiento. De este modo, se pueden analizar los datos, detectar visualmente los espurios o realizar comparaciones entre atributos.

WEKA explorer se ha usado principalmente para realizar el *Benchmarking* en la selección de clasificador y atributos. La herramienta “*Visualize*” de *WEKA* ha sido de especial utilidad para realizar comparaciones dos a dos de los datos, detectando así correlación en ellos (ver Fig. B. 3). Además, ofrece ventajas gráficas como la visualización de árboles de decisión creados, como se observa en la Fig. B. 4.

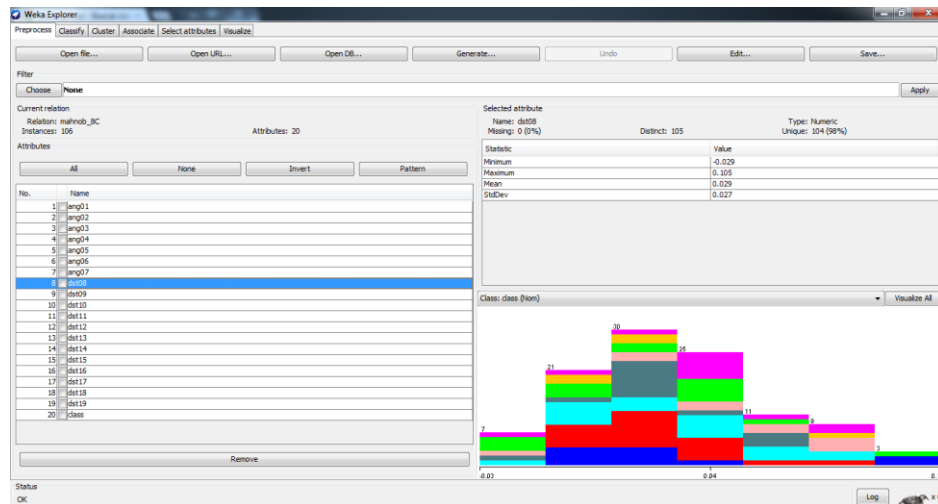


FIG. B. 2. INTERFAZ DE WEKA EXPLORER

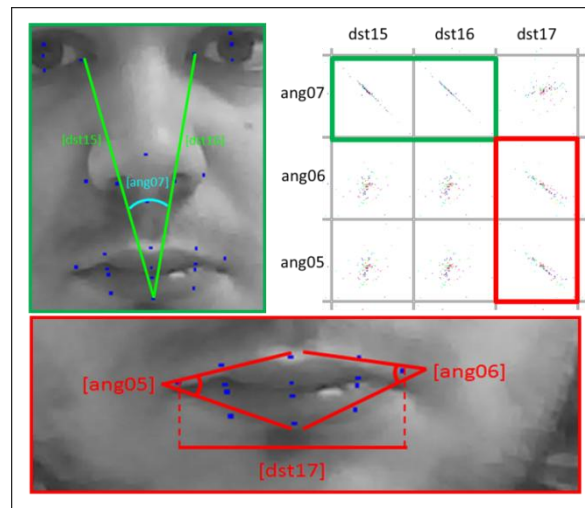


FIG. B. 3. GRÁFICAS DE CORRELACIÓN. EN LA PARTE SUPERIOR LA CORRELACIÓN ENTRE LAS DISTANCIAS 15 Y 16 CON EL ÁNGULO 7 (NARIZ). EN LA PARTE INFERIOR LOS ÁNGULOS 5 Y 6 CON LA DISTANCIA 17 (LABIOS)

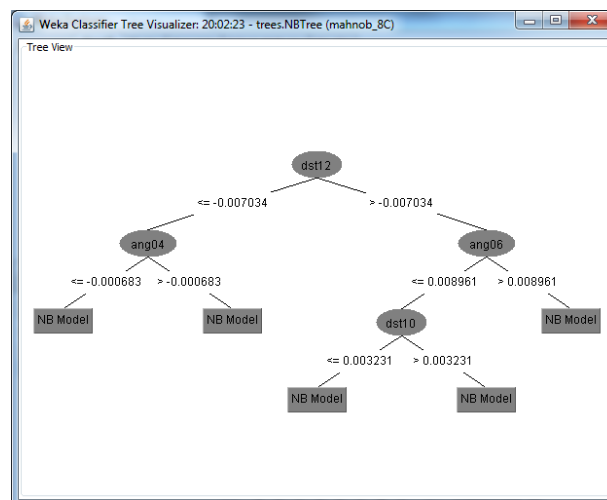


FIG. B. 4. VISUALIZACIÓN DE UN ÁRBOL DE DECISIÓN GENERADO EN UNA SESIÓN DE ENTRENAMIENTO

B.2.2 Librería WEKA para IDE Eclipse

WEKA Explorer permite realizar entrenamientos y clasificaciones mediante su interfaz gráfica, pero para realizar estas tareas de manera automática deben utilizarse las librerías Java que proporciona en un programa creado en el mismo lenguaje.

Las funciones ofrecidas por las librerías de *WEKA* son las mismas que se pueden usar en la interfaz gráfica, y existe una documentación que explica detalladamente el uso de éstas [WDoc].

Para utilizar dichas librerías, se ha integrado el archivo *JAR* disponible en la página web de *WEKA* en el entorno de desarrollo Eclipse para así poder realizar las tareas de clasificación asociadas a la fase 3 del sistema de reconocimiento automático de emociones.

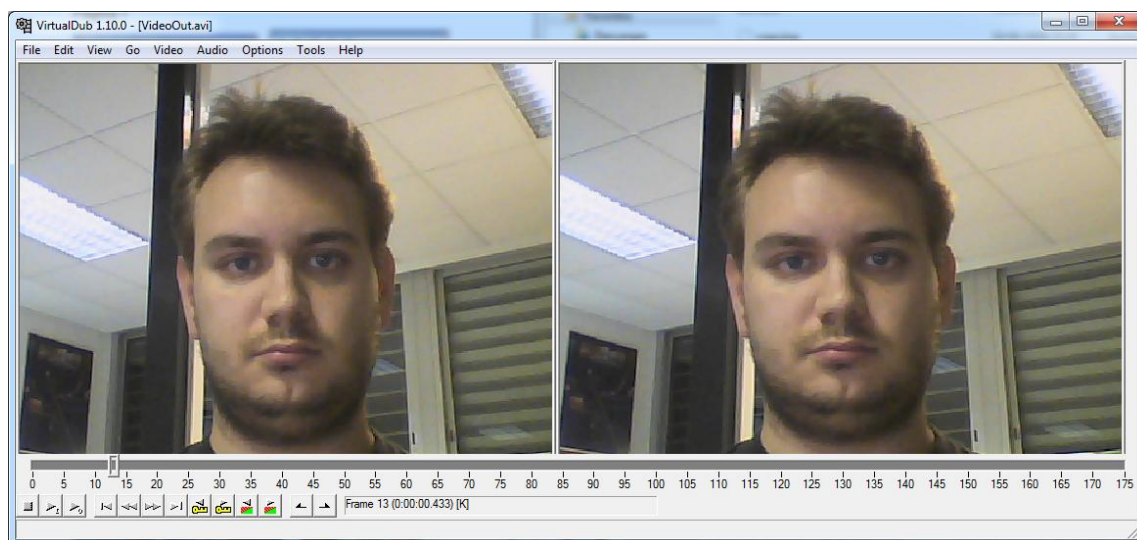
B.3 Captura y Procesado de Vídeo: *VirtualDub*

Para capturar y procesar vídeo existen varias aplicaciones, algunas libres y multiplataforma como *Avidemux*, *VirtualDub* o *Mplayer*. También existen aplicaciones más potentes como *Sony Vegas* o *Adobe Premiere*, todas para Windows. Estas últimas no fueron tenidas en cuenta por no ofrecer ventajas reales para este proyecto. De entre las aplicaciones libres, se seleccionó finalmente *VirtualDub* por su facilidad de uso y ofrecer más funcionalidades que las restantes.

VirtualDub [VDUB] es una aplicación de código abierto desarrollada por Avery Lee para Windows, bajo la *GNU General Public License (GPL)*, que permite la captura y el procesado de vídeo. Es capaz de importar y exportar prácticamente cualquier formato de vídeo, siempre y cuando sus *codecs* estén instalados en el sistema.

La función más útil que ofrece para este PFC, es el recorte de sub-secuencias de vídeo, exportación e importación de fotogramas, conversión entre formatos y visualización y modificación de parámetros de vídeo (como los *frames* por segundo FPS o la resolución de salida). Se ha usado para trocear en *frames* los vídeos de la BDE para entrenar el sistema, y los vídeos de los cuales se quería obtener una evaluación emocional.

En la Fig. B. 5 se puede ver el entorno gráfico de la aplicación *VirtualDub*.

FIG. B. 5. INTERFAZ GRÁFICA DE *VIRTUALDUB*

B.4 IDEs Empleados

B.4.1 Qt Creator

Qt Creator es un IDE (*Integrated Development Environment*) multiplataforma creado por *Trolltech* para el desarrollo de aplicaciones con las bibliotecas Qt, requiriendo su versión 4.x. En el presente proyecto se han utilizado las bibliotecas 4.7, por ser las más recientes y las que más prestaciones ofrecen.

Qt Creator ha sido el IDE empleado en todas las fases de reconocimiento automático de afecto, salvo en la fase de clasificación (dado que se ha usado Eclipse, como se verá a continuación). El lenguaje utilizado en este entorno es C++.

La justificación de la utilización de *Qt Creator*/C++ reside en que son el IDE y lenguaje de programación utilizados en el trabajo previo desarrollado en el grupo. Se estudió la posibilidad de cambiar a lenguaje Java y así poder integrar todo en un mismo entorno de programación. Sin embargo, esta decisión entraba en conflicto con el uso de otras herramientas utilizadas en el grupo para capturar otros canales emocionales como el texto, que tenían la restricción de emplear Sistema Operativo Linux. De manera que se ha mantenido cada fase del sistema de reconocimiento en su IDE original con el lenguaje que previamente ha sido empleado, centrándose el presente proyecto en su tarea más importante de sustituir el clasificador discreto por uno con enfoque continuo.

En la Fig. B. 6 puede observarse el entorno de desarrollo *Qt Creator*.

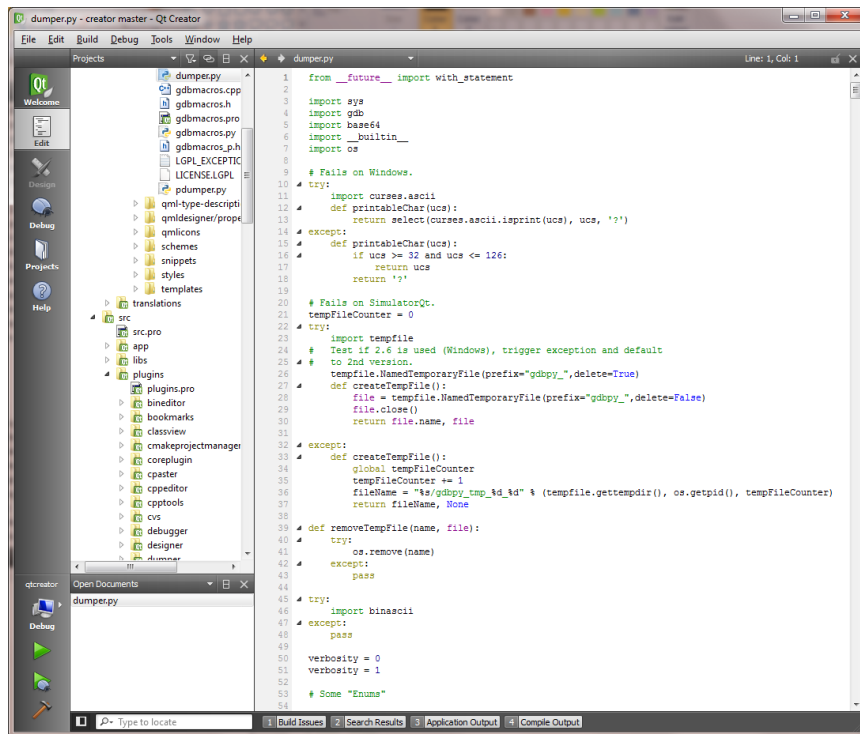


FIG. B. 6. INTERFAZ GRÁFICA DEL IDE QT CREATOR

B.4.2 Eclipse

Eclipse [EIDE] es un IDE multiplataforma y multiarquitectura que inicialmente fue desarrollado por la compañía IBM y que actualmente está gestionado por la fundación *Eclipse*. Está principalmente pensado para el desarrollo de aplicaciones Java (ver Fig. B. 7).

El entorno permite la adición de nuevos componentes (*plugins*) para hacer posible la implementación de proyectos en otros lenguajes de programación (C/C++, PHP, Python, Ruby, etc.), así como otras herramientas útiles para la gestión de proyectos, como las métricas software o el control de versiones.

El uso de la herramienta Eclipse para este proyecto radica en la necesidad de utilizar la librería *WEKA* para las tareas de clasificación. De igual manera que se usa el software *WEKA* para el entrenamiento del clasificador, una vez que esté entrenado, deberá integrarse en el código del sistema de reconocimiento facial y esta tarea se realizará mediante las funciones que *WEKA* proporciona para Java.

Existen otros programas similares, algunos de ellos son también entornos de desarrollo: *Netbeans* o *Microsoft Visual Studio*. También están aquellos que son editores de texto con resaltado de sintaxis, como *Gedit* o *Notepad++*. La elección de eclipse fue debida principalmente a que era el IDE empleado en el trabajo previo y de esta forma se podía aprovecharse código desarrollado.

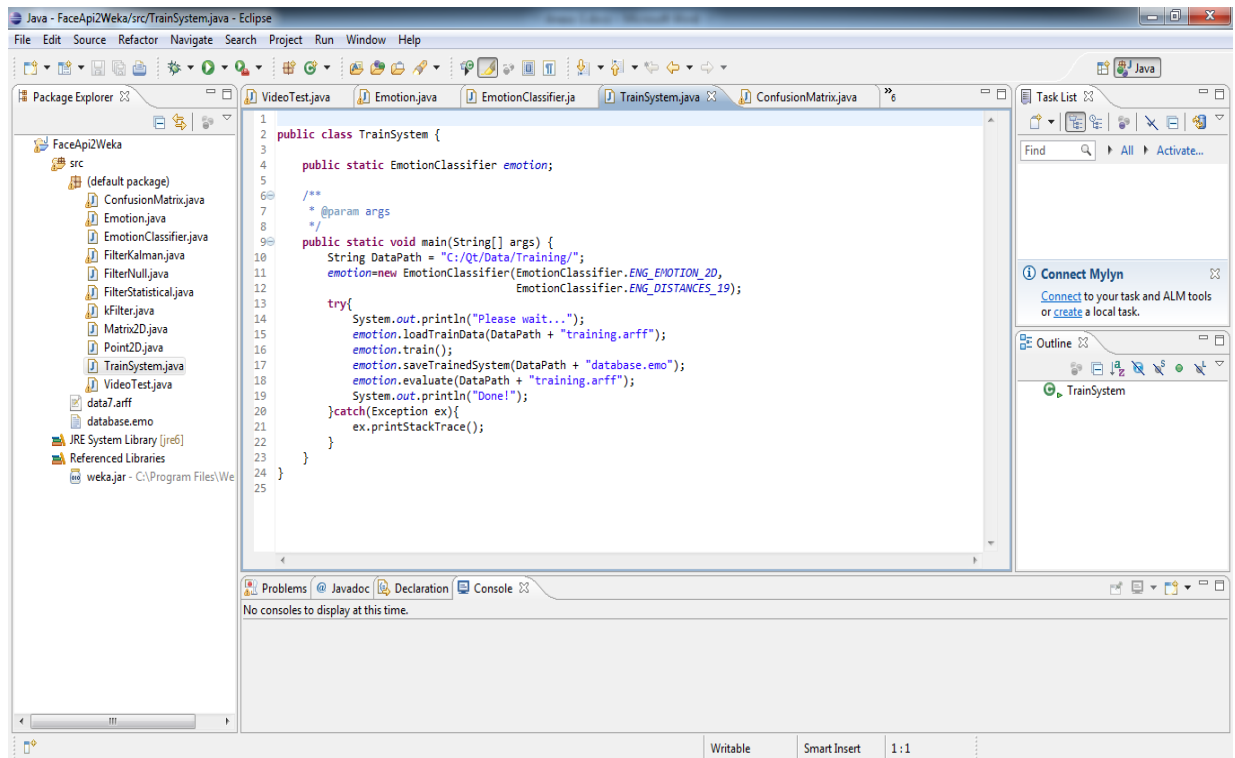


FIG. B. 7. INTERFAZ DEL ENTORNO DE DESARROLLO ECLIPSE

ANEXO C. SELECCIÓN DE SESIONES

Este anexo contiene la información relacionada con la selección de sesiones de la *MAHNOB-HCI Tagging Database*. En la primera parte, se explica la problemática asociada a la selección de datos en la BDE escogida. A continuación, se muestran los datos disponibles, los criterios de selección usados y las sesiones que finalmente han sido escogidas. Por último, se habla de los problemas encontrados durante el proceso de selección de sesiones.

C.1 El problema de la selección de los datos

La Base de Datos Emocional *MAHNOB-HCI Tagging Database* contiene más de 3000 sesiones de sujetos que han sido evaluados emocionalmente por ellos mismos tanto en un enfoque discreto (8 emociones: diversión, enfado, aversión, miedo, alegría, neutral, tristeza y sorpresa.) como en un enfoque continuo (con valores 1-9 para cada una de las 4 dimensiones: evaluación, activación, dominancia y predictibilidad). De todas esas sesiones, 511 están debidamente etiquetadas el enfoque continuo necesario para entrenar al clasificador del sistema de reconocimiento de emociones. Esto es, dichas sesiones contienen todo el material grabado etiquetado con todos los valores.

Como se ha mencionado previamente en la memoria, procesar las 511 sesiones es un trabajo arduo y costoso que excede el tiempo destinado a un Proyecto Fin de Carrera. Por ello, se ha optado por seleccionar solamente parte de todo el material para procesarlo. El problema principal reside seleccionar correctamente aquellos datos que sean útiles para el entrenamiento. Afortunadamente, cada sesión tiene asociado un fichero *XML* que contiene toda la información correspondiente. Concretamente, contiene el etiquetado emocional en continuo, factor que es determinante a la hora de seleccionar sesiones, dado que dicha selección se ha realizado teniendo en cuenta la salida del clasificador: el espacio dimensional en activación-evaluación.

Por tanto, el procedimiento seguido es el siguiente:

- Primero, se extrae toda la información de los ficheros *XML* y se representa mediante gráficas para ver cómo están dispuestas algunas de las características de los datos en función de su etiquetado en activación-evaluación. Es decir, dado un conjunto de sesiones etiquetados en una emoción discreta, se realiza un gráfico de ocurrencias de dichas sesiones en el espacio activación-evaluación, pudiendo observar en qué zonas del espacio dimensional se aglomeran dichas emociones discretas.
- Una vez vista la disposición de los datos, se valoran aquellos aspectos que se deben tener en cuenta a la hora de seleccionarlos, creando así unos criterios de selección como se verá más adelante
- Finalmente, se presentan y se comentan los datos seleccionados.

C.2 Datos disponibles

Una vez analizados los *XML* de las 511 sesiones, se presentan los datos extraídos en una hoja Excel para poder visualizar información en forma de gráficas. Es de especial interés conocer:

- Número de emociones discretas que existen en las 511 sesiones (ver Fig. C. 1).
- Distribución del etiquetado en las 81 coordenadas del espacio continuo activación-evaluación de las 511 sesiones en un mapa de calor (Fig. C. 2).

Como se puede observar en la Fig. C. 1, el número de emociones categóricas en las 511 sesiones disponibles de la BDE escogida no está balanceado. Las emociones de contento (*amusement*), neutral y alegría (*happiness*) son las que más aparecen. Sin embargo, no ocurre lo mismo con emociones como sorpresa (*surprise*) o enfado (*anger*).

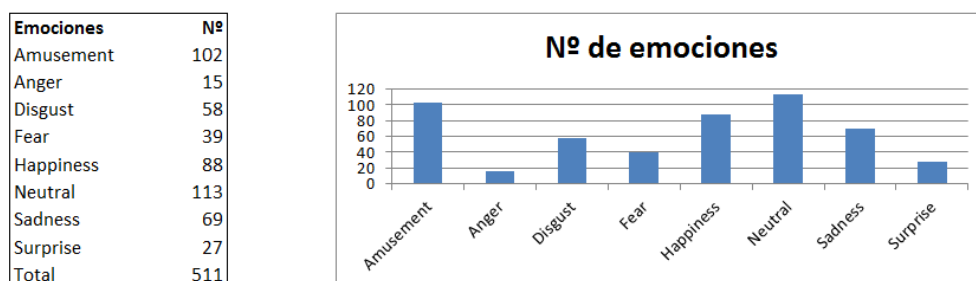


FIG. C. 1. DISTRIBUCIÓN DE LAS EMOCIONES CATEGÓRICAS DE LA BDE MAHNOB-HCI TAGGING DATABASE

En la Fig. C. 2 se observa la distribución de las 511 sesiones en función del etiquetado evaluación-activación. Lejos de existir una distribución uniforme en el espacio de activación-evaluación, las sesiones se encuentran concentradas en la zona media de evaluación, con baja activación. Esto supone un problema dado que este espacio 2D actúa como salida del clasificador, y es necesario que los datos que entrenen al clasificador estén repartidos uniformemente en el espacio 2D activación-evaluación. De lo contrario, el número de instancias por clase no estaría balanceado y el entrenamiento sería incorrecto.

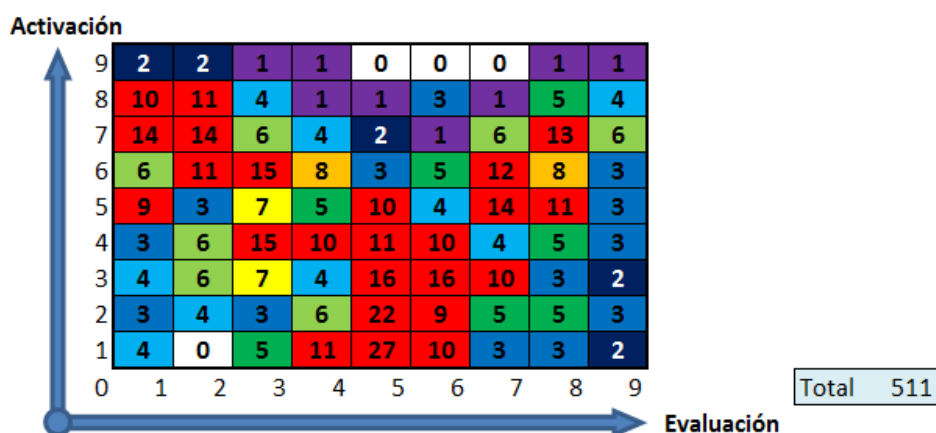


FIG. C. 2. DISTRIBUCIÓN DE LAS 511 SESIONES EN FUNCIÓN DEL ETIQUETADO EN ACTIVACIÓN-EVALUACIÓN

Por tanto, es necesario un balance de datos y por ello se ha optado por elegir un modelo de 9 regiones en lugar de 81, donde las ocurrencias por cuadrante aumentan considerablemente, permitiendo que los datos de la zona superior central se reagrupen y crezcan en número. La Fig. C. 3 contiene el número de ocurrencias en cada una de las 9 regiones de las 511 sesiones.

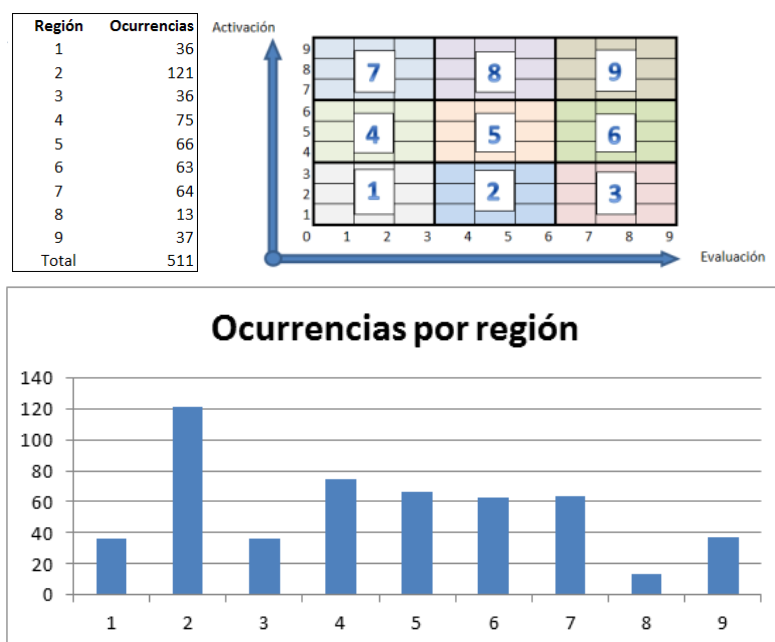


FIG. C. 3. OCURRENCIAS POR REGIONES DE LAS 511 SESIONES

Con el modelo de las 9 regiones que aparece en la Fig. C. 3 se observa que algunas coordenadas que antes estaban vacías de sesiones (e.g las regiones {9,5} ó {9,6} en el espacio activación-evaluación de 81 regiones), han sido ahora agrupadas en la región 8. Sin embargo hay regiones que siguen teniendo pocos datos, especialmente esta última región 8. Aun así, dichos datos serán suficientes para conseguir un balance de regiones equilibrado.

Aunque se podrían agrupar aún más los datos para tener más de éstos en cada una de las regiones, se estaría perdiendo información potencial del espacio continuo en tanto que ya no estaría tan dividido y toda una región correspondería a demasiados posibles valores de etiquetado. Esto es, si agrupamos en 4 regiones (producto de combinar evaluación alta y baja con activación alta y baja), el clasificador solo tendría 4 posibles salidas, resultando ser una clasificación muy pobre y poco continua. Por el contrario, si se intenta agrupar en más de 9 regiones, se estaría volviendo al problema de tener regiones sin apenas datos. Se trata, por tanto, de obtener un compromiso entre el número de regiones y datos disponibles en dichas regiones.

Cabe añadir, y como ha sido comentado en la memoria previamente, que el modelo de las 9 regiones escogido tiene en cuenta solamente 8 de ellas, descartando la región 3. De modo que finalmente las regiones de salida del clasificador son las que aparecen en la Fig. C. 4.

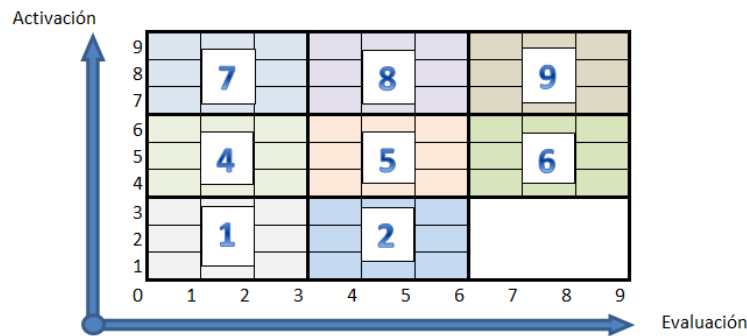


FIG. C. 4. MODELO DE 8-REGIONES ESCOGIDO COMO SALIDA DE CLASIFICACIÓN

Una vez presentados los datos generales y sabiendo que se utilizará el modelo de 8 regiones, el siguiente punto trata sobre los criterios de selección de las sesiones a considerar para el entrenamiento del clasificador.

C.3 Criterios de selección

Por criterios de selección se entienden aquellos aspectos a tener en cuenta a la hora de determinar aquellas sesiones que son más significativas de la BDE escogida. A fin de conseguir unos datos de entrenamiento que traten de englobar todas las características de interés de la *MAHNOB-HCI Tagging Database*, se realizará una selección minuciosa para balancear cada uno de los siguientes aspectos en las regiones de salida:

1. **Nº de datos (instancias).** Cada sesión tiene, entre otras, un par de etiquetas emocionales activación-evaluación que la sitúan como salida en una única región del espacio 2D. Se tratará de conseguir que el número de ocurrencias por región sea uniforme.
2. **Nº de sujetos.** Los datos de entrenamiento deben contener a todos los sujetos de la Base de Datos Emocional distribuidos lo más uniformemente posible en las regiones para no sesgar el entrenamiento a una persona específica.
3. **Nº de emociones categóricas.** Pese a ser una BDE etiquetada en continuo, también está etiquetada en discreto con 8 categorías: diversión, enfado, aversión, miedo, alegría, neutral, tristeza y sorpresa. Aunque la principal aportación del sistema es la migración completa al continuo, no debe obviarse este etiquetado categórico, en tanto que ofrece una relación entre la emoción y la zona del espacio 2D donde ésta se encuentra (como se verá posteriormente en las gráficas). De modo que se tratará de ajustar también la emoción categórica sentida por los usuarios.

Para aplicar estos criterios, se presentan a continuación una serie de gráficas que muestran el balance de sujetos, datos y emociones categóricas.

C.3.1 Balance de número de datos

El balance de número de datos por región consiste en equilibrar el número de sesiones en cada región del espacio de salida, como puede verse en la Fig. C. 3 expuesta anteriormente. En la región 8 se observa que se tienen muy pocos datos (tan sólo 13), por lo tanto si se quiere

balancear el número de datos por región, se debe tomar un número que no sea muy superior a éste en las otras regiones, resultando el total de sesiones en un número bastante inferior a las 511 sesiones disponibles.

C.3.2 Balance de sujetos

Afortunadamente, la aparición de cada uno de los 30 sujetos está uniformemente distribuida a lo largo de las 511 sesiones, como muestra la Fig. C. 5. La mayoría de los sujetos aparecen en un número de sesiones entre 15 y 20 (a excepción de los sujetos nº 12 y nº 15 que no aparecen en ninguna sesión). Asimismo todos ellos están etiquetados uniformemente en las emociones categóricas, de modo que no existe ningún problema serio de balance de sujetos por emoción discreta

La gráfica que corresponde al balance de sujetos por región es muy densa, debido que a se necesitan, para cada una de las regiones, el porcentaje de aparición de cada uno de los 30 sujetos, imposibilitando ver la información adecuadamente. Sin embargo, salvo la región 8 que contiene pocos datos, las demás regiones contienen variedad de sujetos y puede observarse su distribución uniforme en las regiones.

Por tanto, el criterio de selección de sujetos consiste en equilibrar los sujetos que se toman por región, de manera que a la hora de seleccionar las sesiones en cada región, se tenga en cuenta este aspecto y se seleccione un conjunto de sesiones en las que aparezcan la mayor variedad de sujetos distintos. En caso contrario, podría ocurrir que una región solo tuviera datos de un sujeto en cuestión.

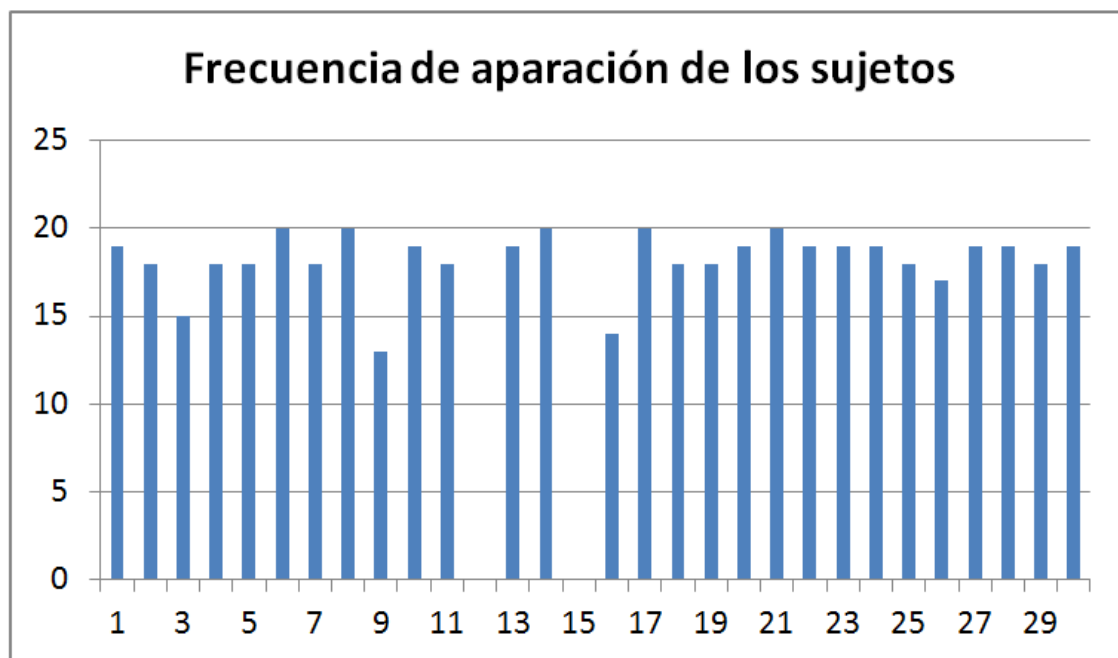


FIG. C. 5. FRECUENCIA DE APARICIÓN DE LOS SUJETOS EN LAS 511 SESIONES

C.3.3 Balance de emociones categóricas

Por último, las sesiones de la *MAHNOB-HCI Tagging Database* están etiquetadas también en una de las 8 emociones categóricas, a saber: diversión, enfado, aversión, miedo, alegría, neutral, tristeza y sorpresa. Asimismo, según el trabajo de investigación de Whissell [Whi89], cada zona del espacio 2D activación-evaluación está asociada a un determinado tipo de emociones, de modo que emociones como alegría o diversión tienen un alto nivel de evaluación (cómo de tan buena o mala es una emoción), y algunas emociones como sorpresa o tristeza son opuestas en el eje de activación (alto o bajo nivel de excitación de la emoción). Por tanto, en las sesiones que finalmente se seleccionen se intentará que la etiqueta emocional categórica se corresponda a la etiqueta continua, evitando de esta manera seleccionar sesiones con etiquetas espurias en las que los sujetos hayan podido cometer errores de auto-etiquetado emocional (e.g. evaluarse en muy baja activación y evaluación si se ha puesto que se siente alegría o diversión).

En la Fig. C. 6 se presentan los mapas de calor que reflejan la distribución de las sesiones etiquetadas para cada emoción categórica en el espacio 2D de activación-evaluación. Los colores simbolizan el número de ocurrencias, que irán desde menor número (colores fríos) hasta mayor número (colores cálidos). Se puede comprobar que, efectivamente, cada emoción categórica se concentra en una zona determinada del espacio 2D, siendo esta asociación muy clara en emociones como aversión o tristeza (baja evaluación), aunque no tan clara en algunos casos como el de sorpresa (más dispersa a lo largo del espacio continuo).

C.4 Datos seleccionados

Finalmente, y tras tener en cuenta los criterios de selección previamente expuestos, se procesaron 250 de las 511 sesiones. Sin embargo, de las 250 sesiones procesadas, algunas no servían como datos de entrenamiento debido a problemas como la mala adecuación de la etiqueta con la expresión facial mostrada, o la imposibilidad del reconocedor facial de reconocer el rostro adecuadamente. Finalmente, se seleccionaron 106 sesiones para el entrenamiento.

En la Fig. C. 7 se presenta la información asociada a las 106 sesiones seleccionadas. En la parte superior izquierda de la figura se encuentra el espacio 2D con las regiones de salida, y la distribución de las emociones categóricas en las regiones. Nótese cómo se han tenido en cuenta los mapas de calor de la Fig. C. 6 para asociar emociones categóricas con regiones. En la parte superior derecha de la figura se observa la distribución de datos por regiones. Finalmente se ha conseguido un balance en el número de datos por región, si bien es cierto que las regiones con menos datos totales como la 8, se han visto afectadas por este hecho. En la gráfica también se puede encontrar un diagrama de tarta en el que se presentan los porcentajes de las emociones categóricas, también balanceadas. Finalmente, en la parte inferior aparece un diagrama de barras que representa la frecuencia de aparición de los 30 sujetos en las seis sesiones de vídeo. Se observa una tendencia que indica que los sujetos con menor índice (1-10) aparecen más veces que los sujetos con mayor índice (11-30). Esto es debido a que los primeros 10 sujetos de la *MAHNOB-HCI Tagging Database* eran idóneos para expresar sus emociones facialmente, mientras que los sujetos que estaban últimos en el orden de numeración no eran tan expresivos.

No obstante, se ha tenido en cuenta a todos ellos intentando balancear el número de apariciones en las sesiones de cada uno de ellos.

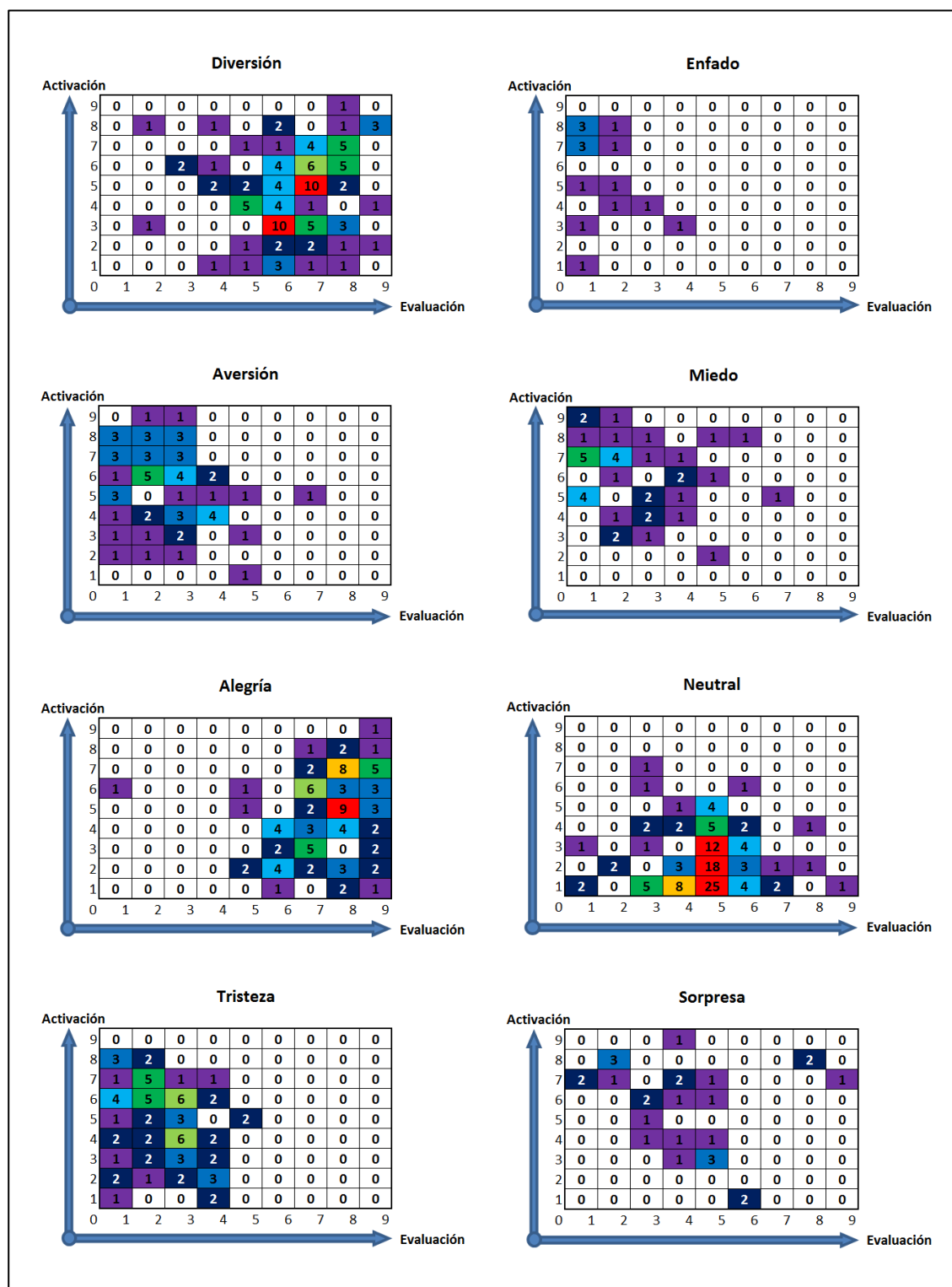


FIG. C. 6. MAPAS DE CALOR DE LAS 8 EMOCIONES DISCRETAS DE LA BDE SELECCIONADA. DISTRIBUCIÓN EN EL ESPACIO 2D ACTIVACIÓN-EVALUACIÓN

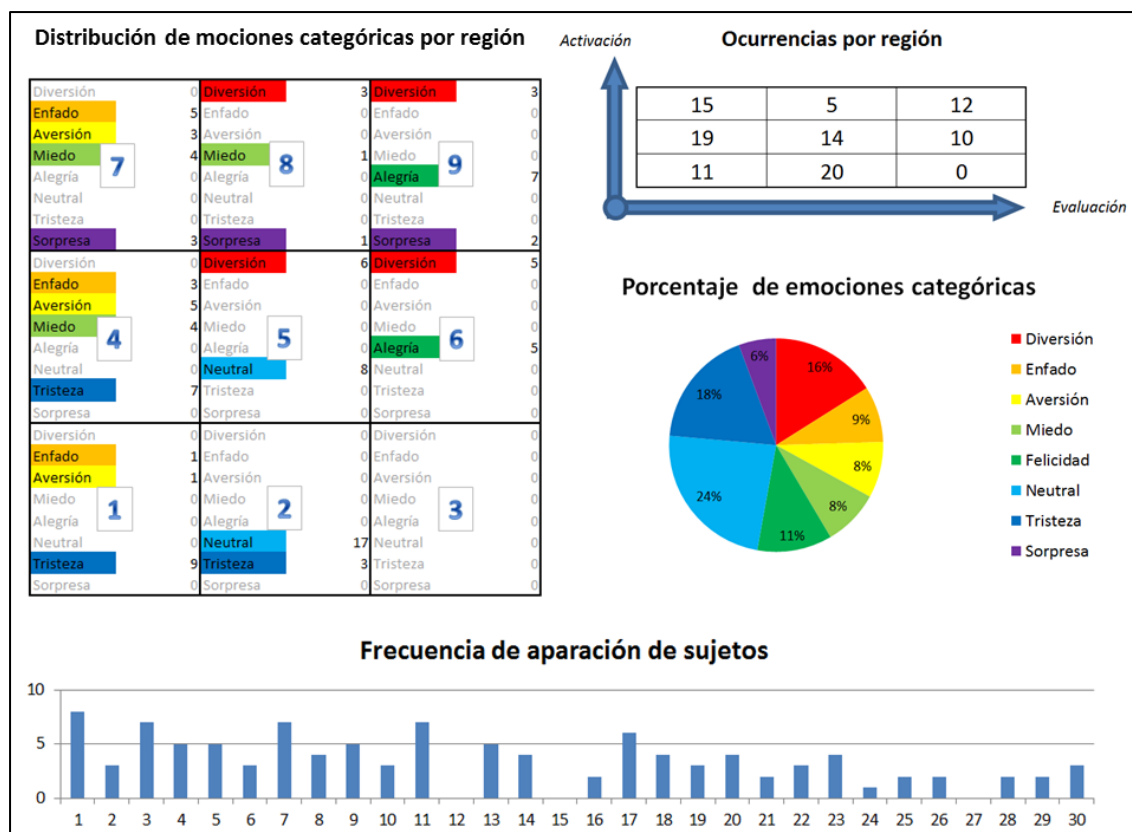


FIG. C. 7. INFORMACIÓN SOBRE LAS 106 SESIONES SELECCIONADAS

C.5 Problemas encontrados

A continuación se procede a explicar los 6 problemas más importantes encontrados durante la fase de selección y extracción de datos de la *MAHNOB-HCI Tagging Database*.

1. Precisión del reconocedor facial

El reconocedor facial empleado ha sido *FaceAPI* de *SeeingMachines*. Un software comercial que permite realizar un seguimiento del rostro en tiempo real proporcionando 38 puntos faciales de interés. Si bien es cierto que el software detecta muy bien la posición de la cabeza y la dirección en la que mira el sujeto, en algunas ocasiones no detecta correctamente los puntos faciales, implicando un reconocimiento erróneo en toda la secuencia de imágenes que le siguen. Esto supone un cuello de botella en la fase de extracción de características faciales, en tanto que se están limitando las salidas a aquellas que ofrezca el reconocedor facial. De este modo, si los datos capturados son correctos, se puede trabajar obteniendo resultados fiables; pero cuando la captura es no satisfactoria, los resultados emocionales obtenidos pueden no ajustarse a los esperados.

2. Distribución desbalanceada en el espacio activación-evaluación.

Como se ha comentado previamente, los datos de la BDE escogida no están distribuidos equitativamente en el espacio continuo activación-evaluación. La solución a este problema, ha sido agrupar las 81 coordenadas del espacio continuo en 9 zonas para disponer de suficientes datos en cada zona y así poder realizar un entrenamiento correcto y balanceado.

3. Expresión no neutral al comienzo de los vídeos

En la fase de división del vídeo en *frames*, es imprescindible que los primeros segundos el sujeto muestre su cara neutra. Esto es debido a que se necesitan esos “datos neutros” para la toma de datos del *baseline*, que permitirá calcular variaciones de distancias en los siguientes *frames* y la normalización los datos.

Normalmente, las Bases de Datos Emocionales faciales cuentan con ello y ponen como requisito a los sujetos de evaluación que mantengan una expresión neutra en el rostro durante los primeros segundos. Sin embargo, y debido a que los sujetos realizan evaluaciones que eliciten emociones diferentes de una manera seguida (más aún cuando los sujetos llevan equipados sensores fisiológicos que requieren de mucho tiempo de preparación), pueden quedar emociones residuales de sesiones ya realizadas en el rostro (e.g, una sesión comienza con un sujeto de evaluación riendo aún por el estímulo mostrado en la sesión previa).

Asimismo, no sólo el rastro de emociones asociadas a sesiones previas impide el reconocimiento de la expresión neutra, sino que acciones como llevarse la mano a la cara, mover mucho los labios o frotarse un ojo, modifican las distancias que se suponen neutras en la expresión, invalidando todos el *baseline* reconocidos.

Para solucionar este problema, durante la fase de división en *frames* del vídeo, se ha asegurado que los primeros segundos (primeros *frames*) contienen una expresión neutra del sujeto de evaluación. En caso contrario, se ha procedido a recortar el vídeo hasta que se consiga el objetivo buscado.

La Fig. C. 8 muestra diferentes casos en los que se impedía el reconocimiento automático de la expresión neutra facial.

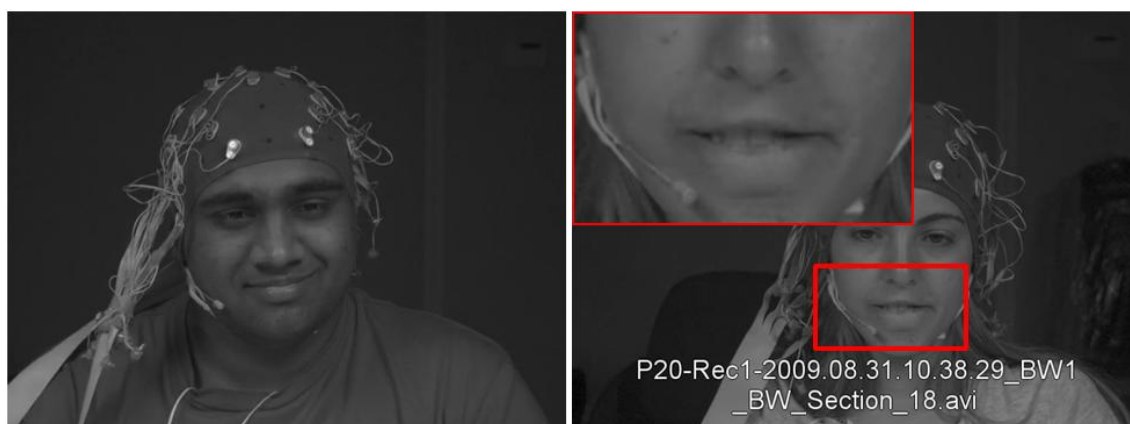


FIG. C. 8. PRIMEROS FRAMES DE DOS SESIONES DIFERENTES. A LA IZQUIERDA, SUJETO Nº3 COMENZANDO EL VÍDEO CON UNA SONRISA. A LA DERECHA, SUJETO Nº 20 COMENZANDO EL VÍDEO MORDIÉNDOSE UN LABIO

4. Sujetos de origen oriental muy poco expresivos

Durante el procesado de sesiones, se observaba que en algunos de los sujetos era muy difícil (a veces incluso imposible) distinguir el momento más fuerte de la emoción, llamado *appex*.

Es necesario determinar el *appex* para identificar el *frame* que representa el estado emocional de la sesión completa.

Este problema ocurría con mayor frecuencia en aquellos sujetos con rasgos asiáticos, que se etiquetaban en zonas de muy alta activación y evaluación siendo las variaciones en sus rostros obtenidas por el reconocedor automático mínimas, como muestra la Fig. C. 9, donde se presenta el *appex* de una emoción etiquetada en una región de máxima activación y evaluación, y sin embargo, se observa que el sujeto mantiene una expresión neutra y calmada sin rastro de emociones.

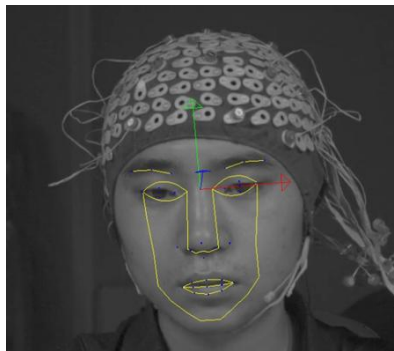


FIG. C. 9. APEX DE UNA SESIÓN ETIQUETDA EN LA REGIÓN 9, CON MÁXIMA ACTIVACIÓN Y MÁXIMA EVALUACIÓN.

Muchas de estas sesiones han tenido que ser descartadas debido a que, con toda seguridad, el clasificador se entrenará con datos erróneos que entrenan emociones como alegría y sorpresa con variaciones que son claramente neutras. Esta es la razón por la cual algunos sujetos aparecen más que otros en los 106 datos seleccionados para el entrenamiento.

5. Expresiones muy rápidas

Durante la selección del *appex* en los vídeos, resultó especialmente difícil seleccionar los *frames* representativos del estado emocional para expresiones en la región superior izquierda del espacio 2D activación-evaluación. La razón es que las emociones asociadas a un valor alto de activación y bajo de evaluación son el miedo, la aversión y el enfado. Por citar un ejemplo, en algunas de las sesiones donde los sujetos sentían aversión o asco, la duración de la expresión facial era muy pequeña, resultando en expresiones muy cortas e intensas. Para asegurar la captura de dichos *frames* sin perder detalle, se tuvo que aumentar la frecuencia de *frames* por segundo (*FPS*), tomando un mayor número de imágenes y aumentando por consiguiente el número de *frames* con el *appex* disponible.

La Fig. C. 10 muestra una secuencia de 6 *frames* ordenada perteneciente a la sesión 2880 del sujeto nº 13. La mujer que aparece en las imágenes se etiquetó en valores del 1 al 9 (con un valor de 2 en evaluación, y un valor de 7 en activación, correspondientes a la región 7 en el espacio continuo 2D de salida). Los *frames* capturados en la figura corresponden a 0.5 segundos en el vídeo de la BDE, reflejando el problema de expresiones muy rápidas que se comenta en este punto.



FIG. C. 10. EXPRESIÓN DE MIEDO EN 6 *FRAMES* CORRESPONDIENTES A 0.5 SEGUNDOS DE VÍDEO

6. Problemas de espacio

Finalmente, otro de los mayores problemas encontrados ha resultado ser el almacenamiento de todo el material de la *MAHNOB-HCI Tagging Database* en disco, dado que son aproximadamente unos 280GB, repartidos en más 1200 carpetas y 400.000 archivos. Para solucionar este problema, se utilizó un disco duro externo de 500GB con conexión USB 3.0 para el almacenamiento de la BDE.

ANEXO D. BENCHMARKING DEL CLASIFICADOR

Este anexo recoge toda la información relacionada con el apartado 3.5 de esta memoria: *Entrenamiento de un clasificador continuo*. Se explica en detalle la metodología que se ha seguido para seleccionar el tanto el clasificador de datos como sus entradas.

D.1 Introducción

Se entiende por *benchmarking* el conjunto de pruebas realizadas sobre el fichero de clasificación ARFF para determinar qué clasificadores y qué atributos (entradas) son los más adecuados para crear el sistema clasificador. La herramienta empleada para realizar las tareas de clasificación es el software *WEKA* (*Waikato Environment for Knowledge Analysis*) de la universidad de Waikato, Nueva Zelanda [WEKA] (ver Anexo B para más detalle).

D.2 Selección de clasificadores para las pruebas

En primer lugar se va a dividir el *benchmarking* en función de los clasificadores que se usen. La herramienta *WEKA* proporciona un conjunto de algoritmos de aprendizaje automático para tareas de minería de datos. De esta colección, se han tenido en cuenta 5 clasificadores: RIPPER, Multilayer Perceptron (MLP), Naive Bayes, C4.5 y Clasificador Sensible al Coste con MLP (CSC-MLP). Esta selección se basa en el uso generalizado de estos clasificadores en el reconocimiento de emociones:

- **RIPPER:** *Repeated Incremental Pruning to Produce Error Reduction*. Es un algoritmo basado en reglas. Se tiene en cuenta este clasificador por su simplicidad y su buen rendimiento.
- **MLP:** *Multilayer Perceptron*. Es la red neuronal más usada para tareas de clasificación. Su utilidad radica en la similitud con ciertas redes neuronales biológicas del cerebro humano, lo cual resulta muy útil para este dominio de trabajo.
- **Naive Bayes:** es un clasificador probabilístico basado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales. Es a causa de estas simplificaciones, que se suelen resumir en la hipótesis de independencia entre las variables predictoras, que recibe el apelativo de ingenuo (*naive*). Sin embargo, puede entrenarse muy eficientemente en un ambiente de aprendizaje supervisado.
- **C4.5:** es también un clasificador basado en reglas, pero se usa para generar árboles de decisión. Su buen rendimiento se debe al uso del concepto de entropía en el ámbito de la información para establecer mecanismos de clasificación.
- **CSC:** Clasificador Sensible al Coste. Este método de clasificación tiene en cuenta una matriz de coste asociada dada por el usuario y un clasificador. En dicha matriz el usuario establece los pesos de cada clase para el clasificador internamente pondere las instancias de entrenamiento. Se usará el *Multilayer Perceptron* como clasificador asociado.

D.3 Selección de atributos

Una vez presentados los clasificadores que se van a tener en cuenta, se realizará una selección de atributos de entre las 19 distancias faciales obtenidas a partir de los 38 puntos faciales que proporciona el reconocedor. El objetivo de la selección de atributos es conseguir diferentes ficheros de entrenamiento con el mismo número de datos o instancias, pero cada una de ellas con diferentes atributos. Esto es debido a que no todos los atributos tienen el mismo peso en la clasificación. De este modo, eliminando atributos y quedándose con los más relevantes, se contribuye a una agilización del algoritmo y la consiguiente reducción del tiempo de proceso, eliminando atributos redundantes y mejorando la clasificación.

No existe una selección de atributos óptima: ésta varía en función del clasificador usado. Por lo tanto, se realizará una selección de atributos para cada uno de los cinco clasificadores. Mientras que RIPPER, Naive Bayes y C4.5 pueden usar los métodos de selección de atributos ofrecidos por WEKA, no ocurre así con los clasificadores que usan redes neuronales debido a su naturaleza, que tiene en cuenta todos los atributos para posteriormente ajustar pesos. Ese es el caso del clasificador MLP y de CSC-MLP, de manera que estos dos últimos clasificadores se tratarán de forma diferente.

Selección de atributos para RIPPER, Naive Bayes y C4.5

La selección de atributos en WEKA consta de dos partes:

1. Método de Evaluación: función que determina la calidad del conjunto de atributos para discriminar la clase. Podemos encontrar dos tipos de métodos de evaluación:
 - a. *Wrappers*: utilizan un clasificador específico para medir la calidad del subconjunto de atributos a través de la tasa de error del clasificador. Se utilizará el método *WrapperSubSetEval* que ofrece WEKA.
 - b. Filtros: calculan la correlación de la clase con cada atributo, y eliminan aquellos redundantes que tienen una correlación muy alta. Se utilizará el método *CfsSubSetEval* que ofrece WEKA.
2. Método de Búsqueda: forma de realizar la búsqueda de conjuntos. Si se quiere realizar una evaluación exhaustiva de todos los subconjuntos, aparece un problema combinatorio inabordable en cuanto crece el número de atributos. Debido a este problema, aparecen estas estrategias de búsqueda de una forma más eficiente. Para una mayor agilidad en el desarrollo del benchmarking, se usará el método “primero el mejor” que ofrece la herramienta WEKA. Además, unas primeras pruebas comparativas han demostrado que no existe mucha diferencia entre los atributos seleccionados al usar el método de “primero el mejor” que un método de búsqueda exhaustivo.

Selección de atributos para MLP y CSC-MLP.

Para CSC-MLP se utilizarán diferentes matrices de coste y, para cada una de ellas, se utilizará el algoritmo “*ranker*” de WEKA que ofrece una lista ordenada de atributos por relevancia. Seguidamente se ponderarán esos *rankings* para obtener un nuevo fichero de clasificación ARFF

que contenga los atributos más relevantes de realizar una clasificación sensible al coste con diferentes matrices. MLP no generará nuevos ficheros.

Generación de ficheros ARFF

La Fig. D. 1 refleja el esquema de selección de atributos mediante el cual se conseguirán nuevos ficheros ARFF para los clasificadores RIPPER, Naive Bayes, C4.5 y CSC-MLP. En la figura, los tres primeros clasificadores comenzando por la izquierda siguen el mismo proceso descrito arriba, obteniendo 4 ficheros ARFF diferentes (marcados en amarillo). Por otra parte, y marcado en verde, está el fichero original con el conjunto completo de atributos, donde no se ha seleccionado ningún atributo de entre los 19. Los archivos van numerados del 1 al 4 precedidos por la letra inicial del nombre del clasificador. Los ficheros acabados en 1 y 3 seleccionan atributos utilizando todo el conjunto (FTS o *Full Training Set*), mientras que los ficheros acabados en 2 y 4 utilizan la técnica de validación cruzada de 10 pliegues (CV o *Cross Validation*) para la selección [WCV].

En cuanto a los clasificadores de redes neuronales (MLP y CSC-MLP), el clasificador CSC-MLP obtiene un único fichero ARFF mediante la ponderación de los resultados obtenidos con las diferentes matrices de coste en el proceso de clasificación. No se generarán nuevos archivos ARFF con métodos WEKA para los clasificadores con redes neuronales, de modo que se utilizarán unos ficheros donde la selección de distancias ha sido manual. Concretamente, se utilizarán:

- Fichero D1: Todas las distancias excepto la de normalización.
- Fichero D2: Distancias simétricas del lado derecho de la cara.
- Fichero D3: Distancias simétricas del lado izquierdo de la cara.
- Fichero D4: Distancias que utilizaba el algoritmo de clasificación previo [Ball11].

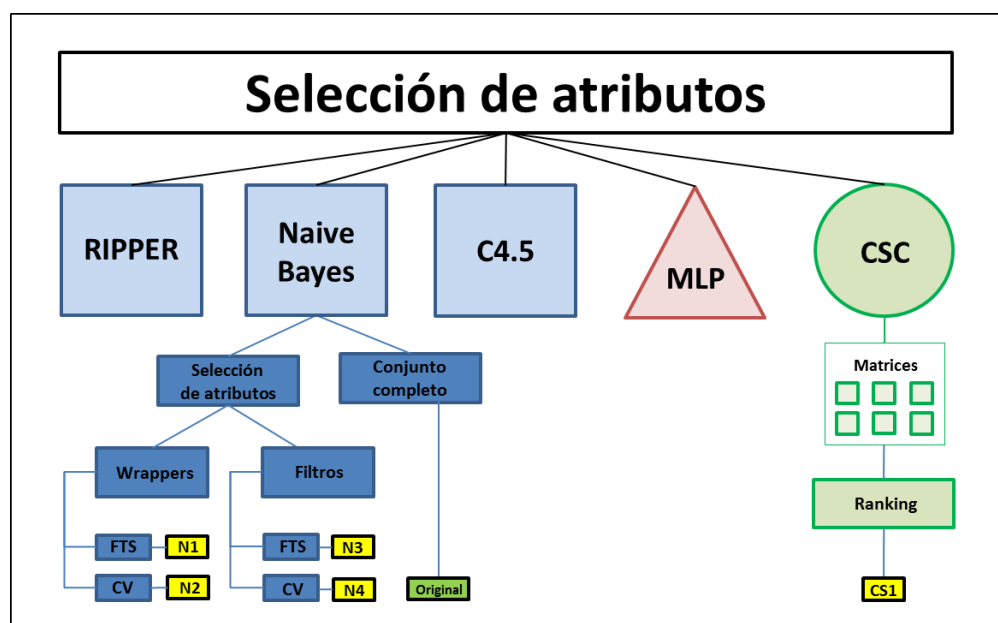


FIG. D. 1. ESQUEMA DE SELECCIÓN DE ATRIBUTOS

D.4 Resultados de *Benchmarking*

Una vez tenemos todos los ficheros ARFF con los atributos seleccionados previamente, cada clasificador entrenará los ficheros que ha generado mediante la selección de atributos (en el caso de los dos clasificadores con redes neuronales, también las distancias seleccionadas manualmente) utilizando dos métodos de validación: validación cruzada de 10 pliegues (*10-fold cross validation*) y conjunto completo de entrenamiento como entrada, obteniendo para cada uno de ellos los resultados. La Tabla D. 1 muestra los resultados de entrenamiento con validación cruzada de 10 pliegues, mientras que la Tabla D. 2 muestra los resultados con el conjunto de entrenamiento entero como entrada.

TABLA D. 1. RESULTADOS DE CLASIFICACIÓN OBTENIDOS MEDIANTE VALIDACIÓN CRUZADA

	Fichero	% Acierto	MLP		Fichero	Descripción
	Original	26%	Fichero	% Acierto	Original	Fichero original con las 19 distancias
RIPPER	R1	28%			Ficheros obtenidos mediante la selección de atributos	
	R2	31%	Original	31%	R1,N1,C1	Fichero resultado de evaluación con wrapper y conjunto completo con los clasificadores Ripper, NB y C4.5 respectivamente
	R3	27%	CSC1	30%	R2,N2,C2	Fichero resultado de evaluación con wrapper y validación cruzada con los clasificadores Ripper, NB y C4.5 respectivamente
	R4	28%	D1	29%	R3,N3,C3	Fichero resultado de evaluación con filtro y conjunto completo con los clasificadores Ripper, NB y C4.5 respectivamente
NAIVE BAYES			D2	28%	R4,N4,C4	Fichero resultado de evaluación con filtro y validación cruzada con los clasificadores Ripper, NB y C4.5 respectivamente
	Original	29%	D3	30%	CSC1	Fichero obtenido ponderando los resultados de selección con métodos ranker para todas las matrices de coste tenidas en cuenta
	N1	34%	D4	30%	Ficheros obtenidos mediante selección manual	
	N2	35%			D1	Fichero con todas las distancias excepto la distancia de normalización
	N3	26%			D2	Fichero en el que se han excluido las distancias simétricas derechas
C4.5	N4	33%			D3	Fichero en el que se han excluido las distancias simétricas izquierdas
			Fichero	% Acierto	D4	Fichero en el que figuran las distancias del clasificador del trabajo previo
	Original	23%	Original	35%		
	C1	27%	CSC1	33%		
	C2	30%	D1	30%		

TABLA D. 2. RESULTADOS DE CLASIFICACIÓN OBTENIDOS MEDIANTE CONJUNTO COMPLETO

	Fichero	% Acierto	MLP		Fichero	Descripción
	Original	46%	Fichero	% Acierto	Original	Fichero original con las 19 distancias
RIPPER	R1	40%			Ficheros obtenidos mediante la selección de atributos	
	R2	43%	Original	72%	R1,N1,C1	Fichero resultado de evaluación con wrapper y conjunto completo con los clasificadores Ripper, NB y C4.5 respectivamente
	R3	42%	CSC1	71%	R2,N2,C2	Fichero resultado de evaluación con wrapper y validación cruzada con los clasificadores Ripper, NB y C4.5 respectivamente
	R4	44%	D1	71%	R3,N3,C3	Fichero resultado de evaluación con filtro y conjunto completo con los clasificadores Ripper, NB y C4.5 respectivamente
NAIVE BAYES			D2	74%	R4,N4,C4	Fichero resultado de evaluación con filtro y validación cruzada con los clasificadores Ripper, NB y C4.5 respectivamente
	Original	47%	D3	60%	CSC1	Fichero obtenido ponderando los resultados de selección con métodos ranker para todas las matrices de coste tenidas en cuenta
	N1	41%	D4	64%	Ficheros obtenidos mediante selección manual	
	N2	42%			D1	Fichero con todas las distancias excepto la distancia de normalización
	N3	40%			D2	Fichero en el que se han excluido las distancias simétricas derechas
C4.5	N4	46%			D3	Fichero en el que se han excluido las distancias simétricas izquierdas
			Fichero	% Acierto	D4	Fichero en el que figuran las distancias del clasificador del trabajo previo
	Original	83%	Original	79%		
	C1	80%	CSC1	82%		
	C2	81%	D1	73%		

Como puede observarse comparando ambas tablas, las tasas de acierto son mucho mayores para las pruebas realizadas sin *cross-validation*. No obstante, se tomarán como válidas las tasas obtenidas mediante la técnica de *10-fold cross-validation* en tanto que ofrecen una visión mucho más realista del comportamiento del clasificador.

A la vista de los resultados obtenidos, se puede decir que los resultados de clasificación están entre un 20% y un 40% de acierto. Se comenta a continuación cuáles son las mejores configuraciones a dos niveles diferentes: método de selección de atributos, modo de selección de atributos.

En cuanto al método de selección de atributos, se ve claramente que los ficheros de atributos generados por *wrappers* ofrecen en los tres clasificadores mejores resultados que los filtros. Esto es debido a que los métodos *wrappers* usan un clasificador a la hora de seleccionar y juegan con ventaja respecto a los filtros, que simplemente usan datos estadísticos para listar los atributos más relevantes. En cuanto al coste computacional, los *wrappers* no son tan rápidos sacando subconjuntos de atributos relevantes. Aun así, y debido a que la selección de atributos se realiza solamente una vez, siempre será preferible usar métodos de selección *wrapper* frente a filtros para futuros entrenamientos.

El modo de selección es análogo a las técnicas usadas en la validación del entrenamiento: se pueden seleccionar atributos utilizando una validación cruzada de 10 pliegues (ficheros R2, N2, C2, R4, N4 y C4), o utilizando el conjunto completo (ficheros R1, N1, C1, R3, N3 y C3). En los tres clasificadores en los que se han usado estos métodos (Ripper, Naive Bayes y C4.5), los resultados para los ficheros de atributos generados con validación cruzada son siempre mejores que los resultados del conjunto completo. Esto es debido a que se han realizado 10 iteraciones en las que, en cada una de ellas, un pequeño subconjunto se ha probado en el total. Asimismo, los ficheros con atributos seleccionados con validación cruzada, además de ser mejores son mucho más fiables, pues garantizan la independencia de datos entre la partición de entrenamiento y prueba.

Clasificador y atributos de entrada seleccionados

Finalmente, en cuanto a la selección del clasificador, se ha optado por elegir el modelo de **Clasificación Sensible al Coste con el clasificador MultiLayer Perceptron** (CSC con MLP) con el conjunto de ficheros generado por éste **CSC1**. La decisión no está solo justificada a que MLP (y por tanto, CSC con MLP) obtengan mejores resultados, ya que se está tratando con diferencias muy pequeñas. La decisión también reside en el hecho de que las redes neuronales tienen una excelente capacidad para extraer información de datos complicados o imprecisos y pueden obtener y detectar patrones que son pasados por alto por humanos u otras técnicas de clasificación. Además, cabe mencionar sus útiles cualidades tales como:

1. Alta adaptabilidad a conjuntos de datos que van cambiando.
2. No realizan ninguna suposición con respecto a funciones de densidad u otra información estadística subyacente en comparación a otros modelos probabilistas [SJC96].
3. Se usan comúnmente en la literatura para reconocimiento de afecto.

La decisión de seleccionar CSC-MLP frente a elegir MLP simplemente, se debe a que se puede condicionar la salida del clasificador en función de las necesidades deseadas a través de una matriz de coste. Esto es especialmente útil si se quiere indicar al clasificador que no importa demasiado si ha clasificado una región como 6 cuando debería ser 9 (ambas situadas en el cuadrante superior derecho del espacio dimensional), pero que es un problema grave que

clasifique un dato en la región 1 cuando debería estar en la 9 (zonas opuestas en el espacio dimensional).

Por lo tanto, pese a que un porcentaje del 33% puede parecer bajo, si se observa la matriz de confusión del clasificador de la Fig. D. 2 (matriz cuadrada de tamaño 8 clases qué indica cómo se han clasificado las instancias) se ve que muchas instancias mal clasificadas que contribuyen a disminuir el porcentaje de acierto, en realidad están clasificando en la zona adyacente. De modo que, aunque solo haya clasificado correctamente el 33% de las instancias, muchos de los fallos no son drásticos y podrían ser tomados como aciertos para algunas emociones que ocupan un amplio rango en el espacio dimensional.

Matriz de confusión del clasificador seleccionado

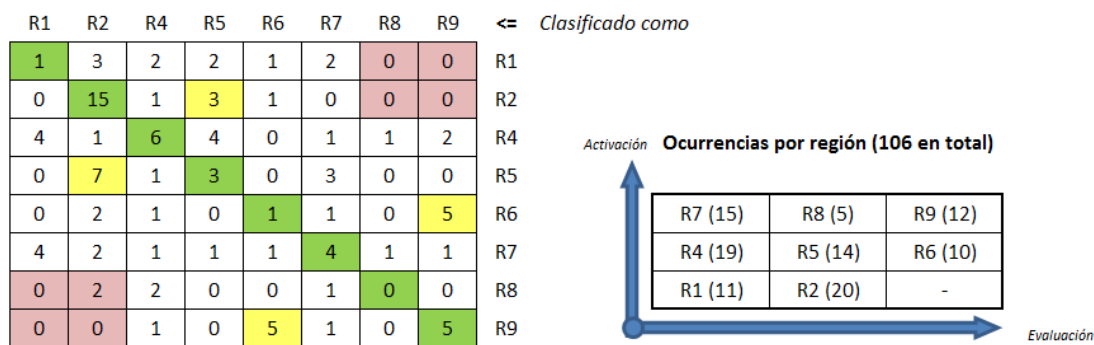


FIG. D. 2. MATRIZ DE CONFUSIÓN DEL CLASIFICADOR CSC-MLP ENTRENADO CON VALIDACIÓN CRUZADA Y EL FICHERO DE ENTRADA CSC-1 CON 12 ATRIBUTOS

ANEXO E. PROCESO DE DESARROLLO DEL SOFTWARE

En este anexo se detalla toda la documentación referente al proceso de desarrollo del software. El primero punto recoge el sistema implementado, donde se exponen las 4 fases del sistema de reconocimiento automático de emociones faciales.

E.1. Sistema de reconocimiento automático de emociones faciales desarrollado

El sistema desarrollado es el que aparece en la Fig. E. 1 Como ya se explicó en el capítulo 3, el sistema consta de 4 fases diferenciadas más otras 2 necesarias para la fase 3 de clasificación.

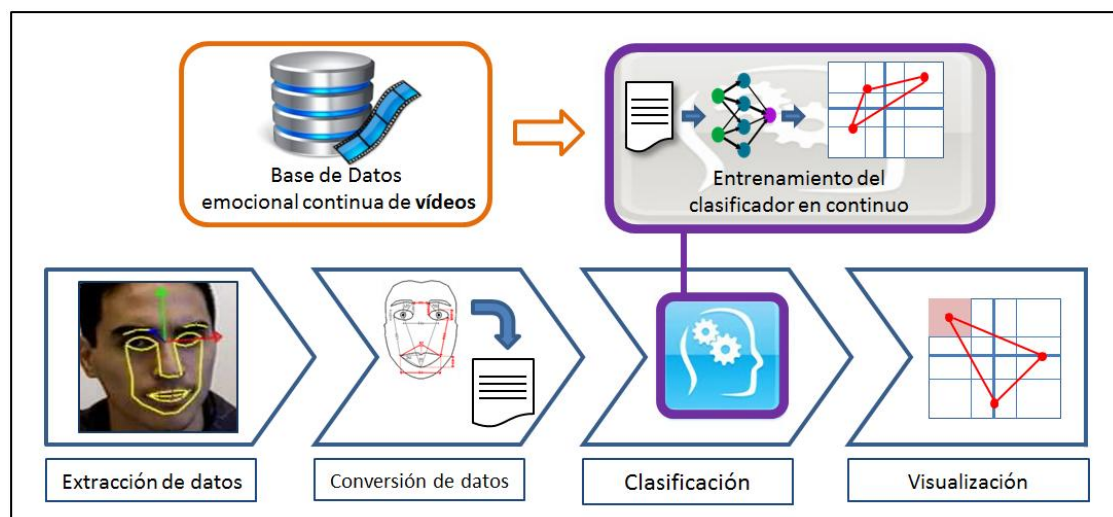


FIG. E. 1. SISTEMA DE RECONOCIMIENTO AUTOMÁTICO DE EMOCIONES FACIALES DESARROLLADO

Para cada una de las 4 fases principales, se incluye el proceso de desarrollo del software seguido. Las fases ubicadas en la parte superior de la Fig. E. 1 (Base de datos emocional de vídeos y entrenamiento del clasificador en continuo) no precisan de desarrollo de software, en tanto que usan programas de edición de vídeo o clasificación ya existentes y las primeras fases del sistema para extracción y conversión de datos. Asimismo, el proceso de desarrollo del software no es homogéneo para las 4 fases, y las fases 1 y 2 no han precisado de éste. Las fases 3 y 4 han seguido un modelo de ciclo de vida incremental, basado en la filosofía de construir incrementando las funcionalidades del programa. La justificación de la elección de este modelo incremental reside en que permite ir aumentando gradualmente las capacidades del software, realizando pruebas tras cada funcionalidad añadida. De esta manera, se sigue un proceso donde se desarrolla independientemente cada funcionalidad, facilitando la labor de desarrollo con la conocida filosofía de trabajo “divide y vencerás”.

E.2 Proceso de desarrollo de la fase 1: extracción de datos

La primera fase del sistema, extracción de datos, consiste en extraer 38 puntos faciales para cada *frame* del vídeo y almacenarlos en un fichero. Esta fase no ha precisado de desarrollo del software, debido a que se ha utilizado íntegramente la fase 1 del trabajo previo [Ball11]. No obstante se ha realizado una exhaustiva documentación del código así como la comprensión de éste, en aras de entender el funcionamiento del sistema completo.

E.3 Proceso de desarrollo de la fase 2: conversión de datos

La fase 2 del sistema de reconocimiento automático de emociones faciales consiste en la conversión de los puntos extraídos en la fase previa a una serie de distancias y ángulos. Esta fase no ha sido analizada, diseñada e implementada desde cero sino que se ha utilizado la fase de conversión de datos del trabajo previo.

Las modificaciones realizadas consisten en cambiar las distancias empleadas en el anterior sistema, por las nuevas distancias y ángulos obtenidos tras la selección de atributos en el apartado 3.5.2 de esta memoria. Las modificaciones han sido realizadas en las siguientes funciones:

4. `int main(int argc, char *argv[])`
 - a. Inclusión de nuevos ficheros de salida y sus correspondientes funciones para escribir los *frames* de entrenamiento.
5. `void writeTrainARFF(QFile *file, QFile *filebool, TFace face, bool header)`
 - a. Modificación de las distancias que se escriben en el fichero de entrenamiento.
6. `void writeARFF(QFile *file, QFile *filebool, TFace face, bool header)`
 - a. Modificación de las distancias que se escriben en el fichero de clasificación.
7. `int readLog(QDir dir)`
 - a. Modificación de la función que lee el fichero generado por la fase 1.
8. `void FaceDST::calcDst (TFace* face)`
 - a. Modificación de la función que calcula las distancias y ángulos a partir de los puntos extraídos en la fase 1.

Asimismo, se ha realizado una completa documentación del código y se han elaborado unas imágenes con los puntos extraídos y distancias y ángulos empleados, que servirán para complementar la documentación y facilitar futuras modificaciones del sistema de reconocimiento automático de emociones faciales.

E.4 Proceso de desarrollo de la fase 3: clasificación

La fase 3 del sistema (clasificación) utiliza un modelo de ciclo de vida incremental, pasando por las fases de Análisis, Diseño, Implementación y Pruebas. De este modo, en la fase de análisis se construirá el Documento de Especificación de Requisitos (DER), donde se incluirán todos los requisitos (funcionales y no funcionales) que implementarán incrementalmente en la fase de implementación, hasta conseguir un programa que satisfaga todos los requisitos.

E.4.1 Análisis de la fase 3

La Tabla E. 1 recoge los requisitos de la fase 3, clasificación.

TABLA E. 1. ANÁLISIS DE REQUISITOS DE LA FASE 3: CLASIFICACIÓN

RF1	Cargar en memoria fichero de clasificación ARFF generado por la fase 2.
RF2	Cargar en memoria fichero de entrenamiento ARFF para entrenar clasificador.
RF3	Crear un clasificador a partir de un fichero de entrenamiento.
RF4	Validar el clasificador entrenado y mostrar porcentajes por pantalla.
RF5	Almacenar el clasificador, así como su matriz de confusión en disco.
RF6	Cargar el modelo de clasificador entrenado en memoria.
RF7	Clasificar cada instancia (<i>frame</i>) del fichero de clasificación ARFF.
RF8	Almacenar las instancias clasificadas en un fichero.
RNF1	Documentar todas las funciones implementadas.
RNF2	Diseñar las estructuras de datos que permitan la escalabilidad del sistema, preparándolas para incluir nuevos clasificadores y matrices de confusión.
RNF3	Hacer el sistema eficiente, de modo que pueda convertirse en un sistema que funcione a tiempo real en un futuro.

Cada uno de los 8 requisitos funcionales corresponde a las funcionalidades que deberá tener el sistema. Los tres requisitos no funcionales no hacen referencia a comportamientos específicos, sino a criterios que pueden usarse para juzgar la operación de un sistema, como puede ser su documentación, su escalabilidad o su eficiencia.

E.4.2 Diseño de la fase 3

Una vez presentados los requisitos del sistema se procede al diseño del diagrama de clases, donde se describe la estructura del sistema mostrando sus clases y relaciones entre ellas. La Fig. E. 2 ilustra el diagrama de clases de la fase 3: clasificación.

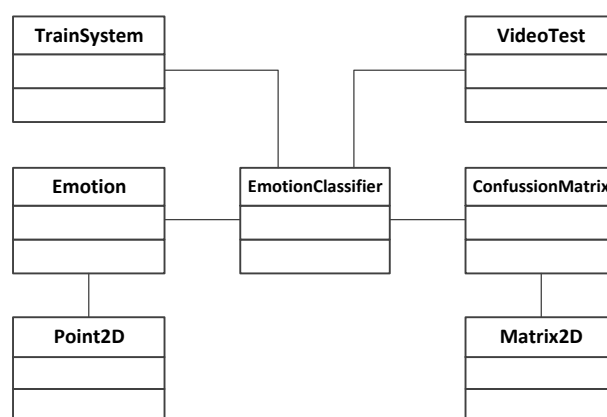


FIG. E. 2. DIAGRAMA DE CLASES DE LA FASE 3: VISUALIZACIÓN

El programa desarrollado en fase 3 contiene dos procedimientos *main()*, ubicados en la clase *TrainSystem* y en la clase *VideoTest*. Esto es debido a que el programa puede usarse de dos maneras diferentes: la primera para entrenar un clasificador a partir de un fichero de

entrenamiento y almacenarlo en disco; y la segunda para clasificar un fichero de clasificación ARFF con un clasificador entrenado previamente.

La elección de un diseño con un programa con dos procedimientos principales se justifica por razones de eficiencia. De este modo, solo es necesario entrenar el clasificador una única vez ejecutando el programa *TrainSystem*. Una vez entrenado el clasificador, se ejecutará el programa *VideoTest* para realizar futuras clasificaciones. En el momento en el que se quiera modificar, incluir o eliminar un clasificador (siempre en un nivel de descripción continuo y con el modelo de 8 regiones) bastará con entrenar de nuevo el clasificador y almacenarlo en disco.

A continuación se explica la función de cada una de las clases en el sistema:

Clase *TrainSystem*:

Permite entrenar un clasificador de entre los disponibles en *WEKA* [WEKA] y almacenarlo en disco para futuros usos.

Clase *VideoTest*:

Permite clasificar un fichero de instancias *ARFF* y convertirlas a un formato compatible con la fase 4 (visualización de resultados).

Clase *EmotionClassifier*:

Clase correspondiente al clasificador emocional. Contiene los métodos de creación y destrucción, así como todos los métodos de entrenamiento y validación, carga y guardado de datos y otras pequeñas funciones.

Clase *Emotion*:

Clase correspondiente a un estado emocional (región) del espacio continuo. Contiene los atributos principales (región del espacio, coordenadas del espacio dimensional) y los métodos *setters* y *getters* (métodos que permiten acceder a los atributos de la clase).

Clase *ConfusionMatrix*:

Clase correspondiente a las matrices de confusión de los clasificadores. Contiene los atributos principales de éstas (filas, columnas, datos) y los métodos *setters* y *getters* correspondientes.

Clase *Matrix2D*:

Clase correspondiente a matrices de dos dimensiones. Contiene los atributos principales (filas, columnas y datos), así como los métodos usuales de operaciones matriciales

Clase *Point2D*:

Clase correspondiente a puntos 2D del espacio. Contiene los atributos principales (valor de x, valor de y), así como los métodos usuales de operaciones entre puntos en dos dimensiones.

E.4.3 Implementación de la fase 3

Una vez realizados el análisis y diseño de la fase 3, se procede con la implementación. Para ello, ha sido necesario implementar todas las clases excepto las clases *Point2D* y *Matrix2D*, que son de libre distribución GNU [WGNU] y se permite la utilización y modificación de éstas.

E.4.4 Pruebas de la fase 3

La fase de pruebas se ha dividido en dos grupos:

- Pruebas de verificación unitarias
- Pruebas de integración

Las pruebas de verificación unitarias consisten en comprobar cada funcionalidad correspondiente a los requisitos funcionales de manera individual, asegurándose de que cada una de ellas funciona correctamente. Las pruebas de integración consisten en probar todo el programa con todas las funcionalidades desarrolladas.

Se han realizado los dos tipos de pruebas, permitiendo descubrir fallos en el código y contribuyendo a la versión final del programa libre de errores y que funciona eficazmente.

E.5 Proceso de desarrollo de la fase 4: visualización

La última fase del sistema (visualización) utiliza un modelo de ciclo de vida incremental, pasando por las fases de Análisis, Diseño, Implementación y Pruebas. Del mismo modo que en la fase 3, se construirá el Documento de Especificación de Requisitos (DER), donde se incluirán todos los requisitos (funcionales y no funcionales) que implementarán incrementalmente en la fase de implementación, hasta conseguir un programa que satisfaga todos los requisitos.

E.5.1 Análisis de la fase 4

La

Tabla E. 2 recoge los requisitos (funcionales y no funcionales) de la fase 4, visualización.

TABLA E. 2. ANÁLISIS DE REQUISITOS DE LA FASE 4: VISUALIZACIÓN

RF1	Cargar los datos de una sesión de evaluación en el visualizador.
RF2	Mostrar en la interfaz el espacio dimensional de 9 regiones y las gráficas de cada uno de los dos ejes en función del tiempo.
RF3	Mostrar en la interfaz los <i>frames</i> del vídeo de evaluación y sus regiones del espacio continuo asociados simultáneamente.
RF4	Recorrer hacia adelante y hacia atrás la secuencia de <i>frames</i> del vídeo mediante botones.
RF5	Recorrer la secuencia de <i>frames</i> mediante un slider correspondiente al tiempo de reproducción del vídeo donde se pueda ir a un frame específico de manera directa.
RF6	Definir un intervalo de <i>frames</i> en el slider
RF7	Mostrar el recorrido emocional del intervalo definido en el espacio continuo.

RF8	Mostrar el recorrido emocional del intervalo definido en las dos gráficas que muestran los dos ejes del espacio continuo vs el tiempo.
RF9	Guardar los <i>frames</i> del recorrido emocional del intervalo seleccionado, así como sus regiones asociadas en disco.
RF10	Guardar la imagen del espacio continuo que aparece en el visualizador en disco
RF11	Guardar los datos (índice del frame, región del espacio) de los <i>frames</i> que componen el intervalo definido en un fichero de texto.
RNF1	Realizar una interfaz intuitiva y de fácil manejo.
RNF2	Las etiquetas de los botones de la interfaz serán en inglés.

Una vez analizados los requerimientos del programa visualizador, se procede con el diseño de éste en el siguiente apartado.

E.5.2 Diseño de la fase 4

Dado que el programa visualizador es una interfaz interactiva, en la primera fase de diseño se realizaron varios prototipos hasta encontrar uno que satisficiera los requisitos. La Fig. E. 3 muestra la interfaz mientras se está visualizando un vídeo, y la Fig. E. 4 muestra la interfaz cuando se está visualizando el recorrido emocional temporal de un intervalo de *frames*.

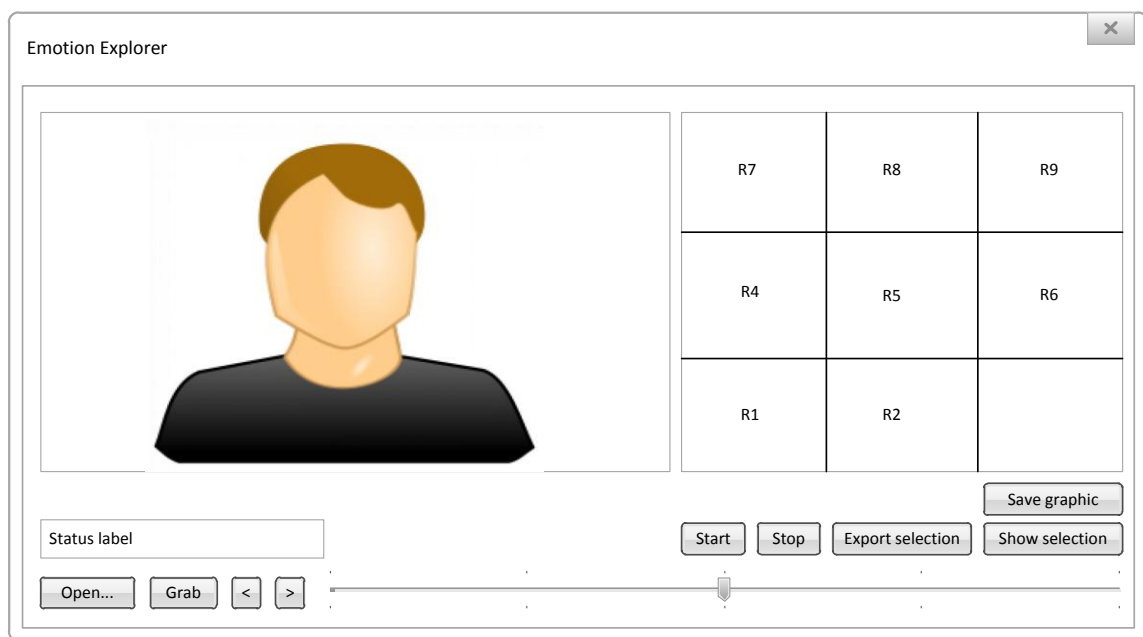


FIG. E. 3. PROTOTIPO DEL VISUALIZADOR DE RESULTADOS MOSTRANDO UN *FRAME* DEL VÍDEO Y EL ESPACIO DIMENSIONAL

Una vez presentados los prototipos, se procede a continuación con el diseño del programa. En este caso, no será necesario un diagrama de clases debido a que todo el programa se puede escribir de una manera clara y organizada en una única clase *player.cpp* que gestiona todas las funciones de la ventana principal.

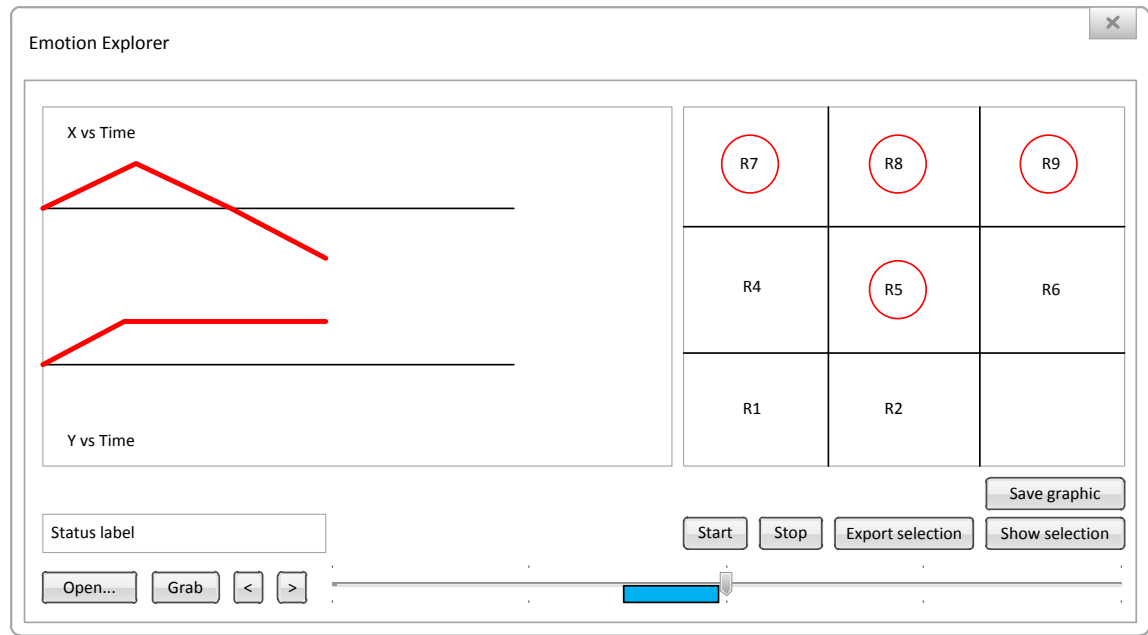


FIG. E. 4. PROTOTIPO DEL VISUALIZADOR DE RESULTADOS MOSTRANDO LA EVOLUCIÓN EMOCIONAL TEMPORAL ASOCIADA A UN CONJUNTO DE *FRAMES* SELECCIONADOS EN EL *SLIDER*

E.5.3 Implementación de la fase 4

La implementación de esta fase se ha realizado en el lenguaje de programación C++ con el IDE *Qt Creator*, como en las fases 1 y 2 del sistema completo. Las librerías gráficas Qt han sido de gran ayuda, contribuyendo a una rápida implementación.

E.5.4 Pruebas de la fase 4

De igual manera que en la fase 3, las pruebas de la fase 4 se ha dividido en dos grupos:

- Pruebas de verificación unitarias
- Pruebas de integración

Gracias a las pruebas de verificación unitarias se ha podido comprobar la correcta funcionalidad de cada uno de los requisitos funcionales iniciales, pudiendo aislar los problemas con el modelo incremental del software empleado. Cabe añadir que gracias al desarrollo de estas pruebas se pudieron detectar y corregir varios errores. Concretamente en el RF6, donde se seleccionaba un intervalo de *frames* con el que posteriormente se podía realizar varias acciones.

Por último, las pruebas de integración consistieron en ejecutar el programa con todas las funcionalidades desarrolladas para comprobar el correcto funcionamiento de éste.

ANEXO F. PUBLICACIÓN INTERACCIÓN 2012

Este anexo contiene el artículo a presentar en el XIII Congreso Internacional de Interacción Persona-Ordenador 2012, que tendrá lugar en Octubre de 2012 en Elche (España).

Proyecto NeuroGame: Neuro-Evaluación Multimodal de Videojuegos

Antonio Loscertales¹, Eva Cerezo¹, Marco Creatura², José Luis González³, Javier Minguez¹,
Rosa Gil³, Sandra Baldassarri¹ and Yolanda López⁴

¹ Departamento de Informática e Ingeniería de Sistemas, Universidad de Zaragoza.
antonio.loscertales@gmail.com, {ecerezo, jminguez, sandra}@unizar.es

² BitBrain Technologies. CEEL, C/ María de Luna nº 11. Nave 4. 50018, Zaragoza.
marcocreatura@bitbrain.es

³ GRIHO, Universitat de Lleida.

jose Luisgs@diei.udl.cat, rgil@griho.net

⁴ Departamento de Psicología y Sociología, Universidad de Zaragoza.
ylopez.iacs@gmail.com

ABSTRACT

Los videojuegos van tomando cada día un papel más importante en el sector del ocio y el entretenimiento, siendo su impacto económico cada vez mayor. Es por ello que la detección temprana, incluso en las fases de diseño, de la aceptación y el impacto del videojuego en los potenciales usuarios se ha vuelto una cuestión de suma importancia. Este trabajo propone una metodología para la evaluación hedónica de videojuegos basada en la medida en laboratorio de las respuestas emocionales y cognitivas de los usuarios. La metodología cubre no sólo la evaluación de la jugabilidad sino de los aspectos más relevantes del diseño que tengan impacto sobre la misma. La metodología se apoya en el uso de las tecnologías actualmente disponibles de detección emocional y cognitiva que permitan la evaluación multimodal a través de diversos canales (visuales, auditivos, fisiológicos). En el montaje experimental propuesto destaca el uso de los sensores fisiológicos que permitan la neuro-evaluación emocional de la experiencia de usuario. La metodología propuesta se aplica en el proyecto NeuroGame a la evaluación intensiva y detallada de tres videojuegos. Se presentan los resultados obtenidos de unas primeras pruebas con usuarios en laboratorio.

Categories and Subject Descriptors

H.1.2 [Models and Principles]: User/Machine Systems – *human factors, human information processing, software psychology*.

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – *evaluation, methodology*.

General Terms

Measurement, Experimentation, Human Factors.

Keywords

Experiencia de Usuario, Evaluación emocional multimodal, Interfaces cerebro-computador, Videojuegos.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Interaccion'12, Oct 3–5, 2012, Elche, Alicante, Spain.
Copyright 2012 ACM 978-1-4503-1314-8/12/10...\$10.00

INTRODUCCIÓN

En el proceso de Interacción Persona-Ordenador, la evaluación de la Experiencia de Usuario (UX) se define como “la percepción de una persona y las respuestas subjetivas de ésta como resultado de la utilización y/o el uso de un producto, sistema o servicio” [ISO/IEC 9241-210, 2010. Clause 2.15], de forma que asume un rol fundamental en el diseño y desarrollo de productos interactivos. En este contexto, la UX y, en particular, la evaluación hedónica (¿qué sentimos al realizar la tarea?) completa las informaciones clásicas más orientadas a una evaluación pragmática (¿para qué sirve?). Este enfoque permite obtener una visión amplia direccionada a mejorar la aceptación del producto, la experiencia interactiva y el retorno económico en fase de desarrollo del producto [6] [11].

Es importante señalar que los estados emocionales usualmente asociados a la experiencia del usuario, influyen procesos cognitivos como la capacidad de atención y memorización, el rendimiento del usuario y, de esta forma, su valoración del producto [16] [2], teniendo así un rol fundamental en la fase de diseño de éste.

El presente artículo presenta NeuroGame, un proyecto multidisciplinar que tiene como objeto profundizar en la comprensión de la medición y el análisis de información asociada a la evaluación hedónica, en forma de respuestas emocionales y cognitivas, para obtener una perspectiva más detallada y precisa de la Experiencia de Usuario. En dicho proyecto participa el AffectiveLab de la Universidad de Zaragoza, el grupo GRIHO de la Universitat de Lleida, la empresa BitBrain Technologies y el Departamento de Psicología y Sociología de la Universidad de Zaragoza. El ámbito de aplicación del proyecto es el del ocio y entretenimiento y, más concretamente, los videojuegos.

El objetivo del artículo es doble. En primer lugar, se propone una metodología general de evaluación de videojuegos tanto en los aspectos de diseño como de jugabilidad, que se centra en la evaluación de la Experiencia de Usuario a través del análisis de las respuestas emocionales y cognitivas de los usuarios medidas en laboratorio. La metodología involucra una serie de tecnologías que permiten la evaluación a través de diversos canales. En segundo lugar, el artículo plantea, basándose en dicha metodología general, la evaluación hedónica de diversos aspectos de tres videojuegos concretos y presenta al final los resultados de las primeras pruebas llevadas a cabo con usuarios.

La estructura del artículo es la siguiente: en el apartado 2 se repasan los trabajos relacionados mientras que en el apartado 3 se detallan las tecnologías involucradas en la metodología presentada en el apartado 4. El apartado 5 detalla la aplicación de la metodología propuesta a la evaluación de la jugabilidad y el diseño de tres videojuegos, presentándose en el apartado 6 los resultados de las primeras pruebas llevadas a cabo. Por último, en la sección 7 se comentan las conclusiones y el trabajo futuro.

TRABAJOS RELACIONADOS

El contexto de los videojuegos y su relación con las emociones es una combinación que en los últimos años ha atraído relativa atención en forma de varios estudios. Concretamente, en el estudio de Ravaja et al. [18] se analizan los patrones de las respuestas emocionales de los usuarios ante varios videojuegos. Los resultados sugieren la utilidad del análisis de los patrones de las respuestas emocionales en el diseño de videojuegos. Otros estudios [14] proponen y estudian la eficacia de las mediciones fisiológicas como evaluación de la UX en el campo de la tecnología de entretenimiento. Los resultados muestran la relación entre las medidas fisiológicas registradas y los cuestionarios proporcionados a los participantes. Este tipo de investigaciones suponen un primer avance hacia la evaluación completa de la Experiencia del Usuario. Recientemente se ha propuesto y llevado a cabo [6] una evaluación de la Experiencia del Usuario respecto a la jugabilidad de un videojuego con técnicas de análisis hedónico. Las conclusiones muestran que factores como la respuesta emocional y la influencia sociocultural enriquecen los métodos tradicionales de evaluación, aconsejando su uso desde etapas más tempranas del desarrollo.

La definición, el modelado y la representación de las emociones sigue siendo objeto de estudio en el campo de la psicología y la computación afectiva [17]. Existen dos enfoques fundamentalmente [7]: el enfoque categórico o discreto y el enfoque dimensional. El enfoque categórico se basa en la clasificación de emociones en categorías [5], por lo que supone una representación discreta de las emociones incapaz de reflejar el amplio rango de emociones complejas que un ser humano es capaz de expresar. El segundo enfoque es dimensional, y establece que las emociones conforman un espacio continuo multidimensional en el que unas y otras no son independientes sino que se relacionan de forma sistemática, siendo el modelo tridimensional (valencia, activación y control) [19] uno de los más extendidos, por lo que ha sido el elegido en NeuroGame y será detallado más adelante.

En cuanto a las respuestas emocionales en sí, estas pueden darse a través de diversos canales. El canal tradicional de detección emocional han sido las imágenes faciales, pero en los últimos años [7] se han ido incorporado otros canales como el sonoro (voz) y, muy especialmente, los fisiológicos (ver Tabla 1). Esto es debido a las crecientes evidencias de que los estados emocionales están relacionados con cambios fisiológicos específicos. En esta línea, se han llevado a cabo estudios experimentales para analizar las respuestas emocionales frente a diferentes tipologías de estímulos: sonidos especiales, música, imágenes, vídeos de películas, videoclips, etc. La Tabla 2 muestra los trabajos más relevantes en esa línea.

A continuación se concreta el despliegue tecnológico propuesto por el proyecto NeuroGame.

TECNOLOGÍAS INVOLUCRADAS

En el proyecto NeuroGame se propone el uso de un despliegue sensorial (ver Tabla 1) que permita recoger información tanto

fisiológica como no fisiológica con el objeto de obtener una visión fiable y completa del estado emocional del usuario. Los sensores fisiológicos se han dividido en aquellos que miden directamente la actividad del Sistema Nervioso Central (SNC) y del Sistema Nervioso Periférico (SNP). En cuanto a los no fisiológicos se persigue básicamente el análisis mediante videocámara de la respuesta facial emocional, complementada por un *EyeTracker* y un micrófono.

Tabla 1. Sensores usados en la literatura para la detección del estado emocional del usuario (en negrita los involucrados en el proyecto NeuroGame).

Sensores Fisiológicos	No fisiológicos
Originados en el Sistema Nervioso Centralizado (SNC): Electroencefalograma (EEG)	Videocámaras (VCF + VCC)
Originados en el Sistema Nervioso Periférico (SNP): Electrocardiograma (ECG) Electromiograma (EMG) Volumen de pulso sanguíneo (BVP) Conductividad de la piel (SC) Respiración (RSP) Saturación de Oxígeno (SpO2) Temperatura de la piel (ST) Electrooculograma (EOG)	Micrófonos (M) EyeTracker (ET)

Sensores fisiológicos

Las señales fisiológicas de origen eléctrico aportan medidas directas e indirectas sobre el estado emocional y además, aquellas que registran directamente la actividad del SNC, pueden aportar medidas sobre funciones cognitivas. A continuación se describe la funcionalidad de cada uno de los involucrados en el proyecto:

- **Electroencefalograma (EEG)**: mide la actividad eléctrica del cerebro. Se realiza colocando un gorro con electrodos puestos en contacto con el cuero cabelludo. Los potenciales registrados son originados en el SNC.
- **Sensor de Volumen de Pulso Sanguíneo (BVP)**: mide de forma no invasiva los cambios relativos de volumen sanguíneo. Esta medición indica la cantidad de sangre que circula actualmente a través de las venas, lo cual permite calcular la vasoconstricción, la dilatación vascular y la frecuencia cardíaca.
- **Sensor de Conductividad de la Piel (SC)**: conocido también como sensor de la actividad eléctrica de la dermis, sirve para medir la capacidad de la piel para la transmisión de corrientes eléctricas, que varían cuando hay sudoración y cambios en el organismo.
- **Sensor de Respiración (RSP)**: mide los cambios de respiración. Con la señal de este sensor se puede calcular parámetros fisiológicos respiratorios como la frecuencia respiratoria, la amplitud relativa de la respiración y la profundidad y rapidez de la ésta.
- **Sensor de Saturación de Oxígeno (SpO2)**: permite monitorizar el nivel de oxigenación de la hemoglobina.
- **Sensor de Temperatura de la Piel (ST)**: sirve para medir la temperatura de la superficie de la piel.

Es importante señalar que no todos los sensores se usan en todas las fases de la evaluación. Esto es debido a que algunos de ellos

Tabla 2. Trabajos relevantes donde se estudian, mediante diferentes sensores de tipo fisiológico, las

Ref.	Estímulos	Sensores	Resultados	Emociones involucradas
Maaoui et Pruski, 2010 [13]	Imágenes	SC, BVP, EMG, ST, RSP	Precisión del 92% en el reconoc. de las seis emociones.	Diversión, alegría, asco, miedo, tristeza, neutro.
Cutbert et al, 2000 [4]	Imágenes	EEG, EMG, ECG, SC	Cambios en parámetros durante estímulos.	Agrado, desagrado y neutro.
Wagner et al, 2005 [20]	Canciones	SC, EMG, ECG, RSP	Precisión en media del 80% en el reconoc. Emocional.	Alegría, ira, placer, tristeza.
Liu et al, 2010 [12]	Sonidos	EEG	Nuevo algoritmo para reconoc. y visualiz. de emociones en tiempo real.	Miedo, frustración, alegría, complacencia, tristeza, satisfacción.
Nasoz et al, 2003 [15]	Fragmentos de películas	SC, ST, ECG	Precisión en media del 76% en el reconoc. emocional.	Tristeza, ira, miedo, sorpresa, frustración, diversión.
Costa et al, 2006 [3]	Fragmentos de películas	EEG	Aumento del índice de sincron. durante estímulos emocionales.	Alegría, tristeza.

tienen un impacto invasivo sobre el sujeto de evaluación por lo que no es posible su uso en algunas fases (e.g. sensores colocados en los dedos no permiten al usuario jugar de forma natural).

Sensores no fisiológicos

Los sensores que miden actividad fisiológica se complementan con otros sensores, los cuales aportan información emocional y de atención visual. Los sensores asociados son:

- **Videocámara frontal (VCF):** pequeña videocámara situada sobre la pantalla, enfrente del usuario, con objeto de grabar las expresiones faciales del mismo. Trabaja a una resolución de 752x480 píxeles con una frecuencia de 60 FPS (fotogramas por segundo).
- **Videocámara de campo (VCC):** se utiliza para monitorizar toda la actividad durante el experimento. Su objetivo es conservar un registro de todo lo ocurrido durante la evaluación. Se empleará una resolución de 1280x720 píxeles y una frecuencia de 30 FPS.
- **Micrófono (M):** se utilizará el micrófono para registrar las palabras o sonidos que el sujeto de evaluación pueda realizar durante el experimento.
- **EyeTracker (ET):** sirve para evaluar el punto donde se fija la mirada. Aunque existen de diversas características, en el proyecto se dispone de una tecnología consistente en unas gafas con una cámara integrada en la parte superior del ojo derecho del usuario.

A diferencia de los sensores fisiológicos, las videocámaras necesitan para su correcta recogida de datos unas determinadas condiciones de iluminación, posición y frecuencia:

- Iluminación: debe alcanzarse un nivel suficiente para que el rostro del usuario sea capturado adecuadamente por la videocámara frontal (VCF) de forma que el seguimiento de las características faciales necesarias para la clasificación emocional sea factible [9].
- Posición: es muy importante, pues de ella depende la calidad de los datos que serán evaluados. De acuerdo con [10], se realizaron varias pruebas previamente con las videocámaras del experimento, y finalmente se concluyó que el sujeto debería colocarse a 1.2 metros de la cámara frontal de modo

que su cara quedara centrada horizontalmente y ligeramente por encima del eje horizontal que divide la imagen en dos.

- Frecuencia: En el experimento se ha utilizado una videocámara frontal especializada en tecnología para visión artificial y aplicaciones de visión por computador. Se utilizará aquella configuración que proporcione la mayor frecuencia posible para garantizar un correcto seguimiento facial.

METODOLOGÍA

A continuación se propone una metodología general de evaluación multimodal y emocional de videojuegos. En este apartado, se definirá el objeto y los sujetos de estudio, las dimensiones emocionales involucradas y el proceso de evaluación a seguir.

Objeto de estudio

Uno o varios videojuegos elegidos por sus similitudes en cuanto género, historia y protagonistas.

Sujetos de estudio

El tamaño muestral del experimento será de unos 20 sujetos balanceando el número de hombres/mujeres en la medida de lo posible. Se reclutarán personas atendiendo a los siguientes criterios de inclusión y exclusión:

Criterios de inclusión: 1.- Edad: 18-34 años. 2.- Ausencia de tratamiento farmacológico o psicológico actual o en el mes previo al estudio. 3.- Correcta comprensión de español hablado y escrito 4.- Aceptación de consentimiento informado de su participación en el estudio.

Criterios de exclusión: 1.- Trastorno psiquiátrico o físico que impidiese el correcto desarrollo del entrenamiento y/o evaluación. 2.- Imposibilidad de acudir o de realizar las sesiones de evaluación. 3.- No aceptación de consentimiento informado para su participación en el estudio. 4.- No haber jugado a ninguno de los videojuegos seleccionados para el experimento.

Para asegurar que el sujeto de evaluación cumple tanto los criterios de inclusión como los de exclusión, deberá rellenar un test antes de ser aceptado como sujeto del experimento. En el test se incluirán todas las preguntas necesarias para asegurar que el candidato cumple con el perfil necesario, además de aquellos aspectos de origen sociocultural, útiles para la relación entre los

resultados y el perfil de usuario, así como un consentimiento informado donde se explique claramente la naturaleza de las pruebas y el rol de los sujetos en ella, que deberá ser firmado por duplicado por el sujeto y un miembro del equipo evaluador.

Dimensiones emocionales

Se opta por un enfoque dimensional basado en el modelo de emociones de Russell [19]. En este modelo la variabilidad de una emoción se representa mediante tres dimensiones: activación, valencia y control o dominancia. La dimensión de activación se refiere al grado de excitación o apatía de la emoción y va desde emociones apáticas como el aburrimiento hasta emociones de máxima activación como la ira o la sorpresa. La dimensión de valencia se refiere a la forma negativa o positiva de la emoción, oscilando desde sentimientos desagradables hasta sentimientos agradables como la felicidad. Por último, la dimensión de control se refiere al grado de dominancia o control de la situación (i.e. si la situación supera al sujeto que está sintiendo la emoción o por el contrario es éste quien controla la situación). Durante el experimento, se les pedirá a los usuarios que realicen pequeños test o valoraciones donde deberán indicar las emociones experimentadas. Para ello, se utilizará la escala SAM [1] (ver Figura 1). En ella, los usuarios deben etiquetarse en las tres dimensiones de interés (activación, valencia y control o dominancia) en una escala numerada del 1 (mínimo) al 9 (máximo). Además, la escala SAM viene acompañada de dibujos que van acordes a la valoración numérica, facilitando la comprensión y etiquetado en los sujetos de evaluación.

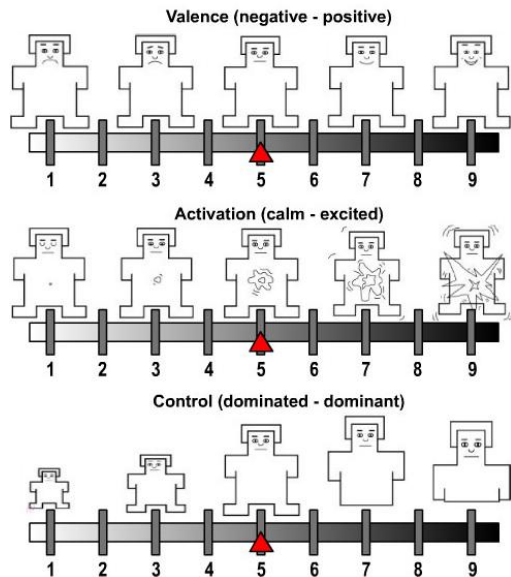


Figura 1. Escala SAM (Self-Assessment Manikin) a usar para la autoevaluación emocional.

Procedimiento

La evaluación se estructurará en una o varias sesiones. Todas las sesiones de evaluación constarán de dos partes bien diferenciadas: calibración y evaluación. En ambas se presentarán al usuario estímulos de una o varias modalidades (imágenes estáticas, videos, melodías,...). El objetivo de la calibración será estimular a la persona en las dos condiciones extremas dentro de cada una de las tres dimensiones emocionales (activación, valencia y control o dominancia). De este modo, se construirán las referencias para la evaluación posterior de la respuesta emocional del usuario. La evaluación consistirá en la presentación de estímulos diana y de

ruido (otros estímulos no relacionados con el tema de interés), obteniendo las medidas sensoriales de respuesta para su posterior análisis. En una evaluación se considerarán una o varias categorías del videojuego (por ejemplo, los personajes o la ambientación) que se considere interesante evaluar. Para cada una de las categorías se presentarán una serie de muestras. Una muestra se compondrá de un descanso, un estímulo (ruido o diana) y una valoración por parte del usuario. En la Tabla 3 se presenta un glosario de términos con objeto de ayudar a la comprensión del proceso.

Tabla 3. Glosario de términos relacionados con la evaluación

Término	Definición
Sesión	Unidad de tiempo de trabajo con el sujeto.
Modalidad	Tipo de estímulo (imagen, melodía, vídeo o interacción).
Calibración	Parte de la sesión dedicada a tomar medidas con el objeto de usarlas como estados de referencia. Una calibración está asociada a una modalidad y se descompone en muestras.
Evaluación	Parte de la sesión dedicada a tomar medidas con objeto de evaluarlas. La evaluación puede ser de una o varias categorías.
Categoría	Tipología de elementos de un videojuego a evaluar en una sesión. Una categoría está asociada a una modalidad y se descompone en muestras.
Muestra	Conjunto de acciones a realizar cuando se presenta un estímulo. Una muestra está siempre compuesta de descanso, estímulo y valoración.

Sesiones

En este apartado se detallan las partes principales de las que se compone una sesión. Como se ha mencionado anteriormente, se denomina sesión a una unidad de tiempo de trabajo con el sujeto. Debido a la duración que supone analizar todas aquellas categorías de interés en un experimento, es muy probable que éste deba partirse en varios días, realizando una sesión por cada uno de ellos. En la Figura 2 se presenta el esquema general de una sesión. A continuación se detallan cada una de las tres partes principales de una sesión: Configuración Inicial, Calibración y Evaluación.

Configuración Inicial

La configuración inicial es imprescindible al comienzo de cada sesión ya que en ella se realizará la puesta a punto de todos los sensores del experimento. Durante esta parte, además de la comprobación del funcionamiento y la configuración de sensores, se realizará también la toma del *baseline* en la que todos los sensores tomarán mediciones en las que el sujeto permanecerá en situación de reposo total. El propósito de la toma de esta medida es la adquisición de un punto de referencia en el que el usuario se encuentra en el estado neutro (sin emociones). No será necesario mostrar al sujeto de evaluación ningún estímulo para esta etapa.

Calibración

La calibración es un paso fundamental para lograr los objetivos, dado que las medidas calibradas se utilizarán para encontrar las relaciones entre las emociones y las señales fisiológicas para el usuario concreto que esté llevando a cabo la evaluación. La información de calibración será útil durante el procesado de datos de modo que puedan conocerse tanto los estados neutros como los límites en las dimensiones emocionales evaluadas para el sujeto calibrado (i.e. un sujeto de evaluación puede mostrar una leve

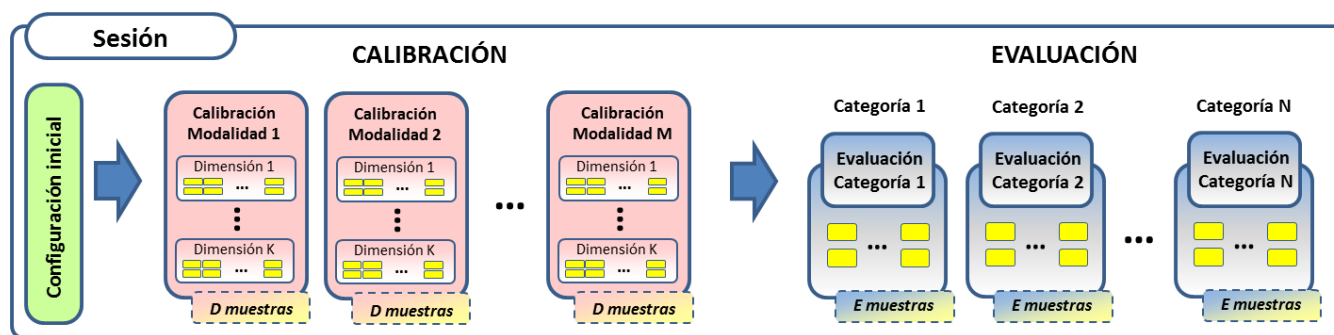


Figura 2. Esquema general de una sesión

sonrisa al ser inducido con un estímulo etiquetado en el extremo de una dimensión, mientras que otro puede mostrarse mucho más expresivo con el mismo estímulo). La calibración puede ser de 4 tipos distintos, atendiendo a las diferentes modalidades: imagen, melodía, vídeo e interacción. A su vez, para cada modalidad habrá que calibrar cada una de las dimensiones emocionales. Por último, por cada dimensión emocional a calibrar se presentarán al sujeto un número de muestras divididas en dos cantidades iguales: la primera mitad de las muestras contendrá estímulos etiquetados en el máximo de esa dimensión, y la segunda mitad etiquetados en el mínimo. El orden de la presentación de las muestras en la calibración de una modalidad para una determinada dimensión será aleatorio. Como se ha visto antes en el glosario las muestras de la calibración contienen, además del estímulo, un pequeño tiempo de descanso al inicio y una valoración del sujeto al final (donde éste realizará un pequeño test y una autoevaluación en las dimensiones que se esté calibrando). Es importante señalar que una calibración en una determinada modalidad debe hacerse siempre que posteriormente se vaya a evaluar categorías asociadas a dicha modalidad. Esto es, si en una sesión van a evaluarse personajes en forma de imágenes, deberá hacerse previa y necesariamente una calibración de imágenes.

Evaluación

La evaluación es aquella parte de la sesión dedicada a tomar medidas con objeto de su posterior análisis. Durante la parte de la evaluación de una sesión, se evalúan categorías (elementos del videojuego). Cada categoría tiene asociada una modalidad. Esto es, si se desea evaluar la categoría relacionada con la banda sonora del videojuego, necesariamente estará asociada a la modalidad melodías. Una categoría está compuesta de un número determinado de muestras. Las muestras, del mismo modo que en la calibración, están compuestas de un breve descanso, un estímulo y una parte de valoración. Los estímulos también estarán divididos a partes iguales, pero en el caso de la Evaluación será en estímulos diana (aquellos con contenidos de los videojuegos a evaluar) y estímulos ruido (aquellos no relacionados con los videojuegos de interés). El descanso será un breve periodo de tiempo durante el cual el sujeto deberá relajarse y la valoración es un pequeño test donde el sujeto responderá preguntas relacionadas con el estímulo y la categoría, autoevaluándose en las dimensiones calibradas.

EVALUACIÓN PROPUESTA

El objetivo final de NeuroGame es llevar a cabo una evaluación multimodal y hedónica comparativa del diseño y la jugabilidad de varios videojuegos, con especial atención a los aspectos socioculturales que tengan relevancia en el diseño del juego. En

este apartado se aplica la metodología general presentada en el apartado anterior al diseño de una evaluación concreta. La evaluación tendrá como objeto el análisis de tres videojuegos: *Castelvania: Lords of Shadow*, *Dante's Inferno* y *God of War III*, elegidos por sus similitudes en la historia y en los protagonistas, por estar disponibles en la misma plataforma de juego (Sony Playstation 3 – PS3) y por ser del mismo género (i.e. *hack and slash* – basados en combates con armas cuerpo a cuerpo).

La evaluación se llevará cabo en tres sesiones, atendiendo a tres grandes aspectos a evaluar en un videojuego. La primera sesión plantea una Evaluación del Diseño, donde se estudiarán las características del diseño sin interacción, analizando por separado las diferentes componentes estéticas que dan forma característica a cada videojuego (es decir, las imágenes de los personajes principales, las melodías, los escenarios y la historia). La segunda fase es una Evaluación de la Interacción, donde se estudia la interacción directa del usuario con el videojuego, evaluando la integración en un entorno dinámico de las características analizadas previamente durante la fase de juego. Por último, se realizará una Evaluación de la Comunicación, en la cual se valora la efectividad de los elementos utilizados en la campaña promocional del videojuego.

A continuación se explican las tres sesiones propuestas detallando las modalidades, categorías y el número de muestras que intervendrán en cada una de ellas. La Figura 3 muestra el proceso completo de la evaluación propuesta.

Evaluación del Diseño

El objetivo de esta evaluación es analizar la respuesta emocional y cognitiva producida por aquellas categorías de un videojuego relacionadas con la experiencia de usuario en la fase de juego. Tras realizar la configuración inicial y la calibración en las modalidades de imagen, melodía y vídeo, se procederá a evaluar las siguientes categorías: 1. Personajes, 2. Contexto, Entorno y Ambiente; 3. Música y Efectos Sonoros; 4. Historia y Narrativa. Los estímulos de las muestras (seis por categoría) le serán presentados al sujeto en las modalidades calibradas: imágenes para las categorías 1 y 2, melodías para la categoría 3 y vídeos para la categoría 4. En la evaluación de la primera categoría, le serán presentadas al usuario imágenes de los protagonistas de los videojuegos objeto de estudio. La segunda categoría se centra en el entorno del videojuego, donde le serán presentados al usuario paisajes y lugares ambientados en los videojuegos. La categoría Música y Efectos Sonoros se centrará en reproducir su música y banda sonora. Por último, la categoría Historia y Narrativa presentará al sujeto la sinopsis o fragmentos de los videojuegos en forma de vídeo.

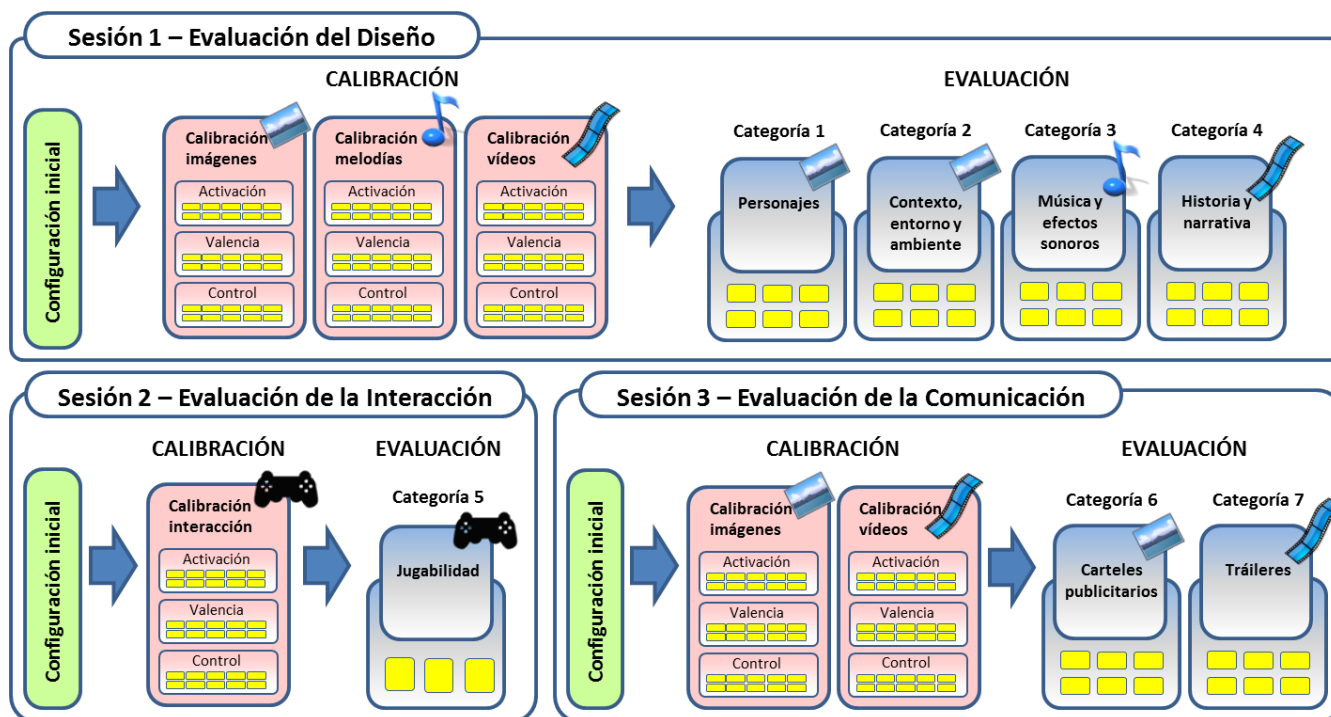


Figura 3. Esquema del proceso completo de la evaluación propuesta

Evaluación de la Interacción

El objetivo de esta fase es evaluar cómo las emociones se relacionan con la estrategia de juego ante los retos propuestos. Se analizará la dinámica de la respuesta emocional y cognitiva del sujeto durante su experiencia directa con el juego. El proceso será el siguiente: primero se realizará la configuración inicial. Dado que la única categoría a evaluar es la Jugabilidad, se calibrará en la modalidad Interacción. Una vez calibrado el sujeto, seguiremos con la evaluación.

La categoría a evaluar estará compuesta por 3 muestras. Cada una de ellas contiene un breve periodo de descanso, un estímulo en forma de una de las demos de videojuego y una parte de valoración en la que se pide al sujeto que evalúe la sección jugada. Durante la demo, el sujeto se enfrentará al videojuego, debiendo superarla con éxito (para ello, todas las demos se configuran en un nivel de dificultad adecuado de forma que puedan ser superadas en un tiempo razonable).

Evaluación de la Comunicación

El objetivo de la última sesión es evaluar la efectividad de los elementos utilizados en la campaña promocional. Para el experimento, el estudio se centra en las categorías de carteles publicitarios y tráileres de los diferentes videojuegos objeto de estudio. El proceso comenzará de manera habitual con la configuración inicial, seguida de la calibración en las modalidades de imágenes y vídeo. La evaluación de ambas categorías se desarrollará presentando al sujeto de evaluación 6 estímulos de imágenes en forma de carteles y pósters para la primera categoría, y 6 estímulos de vídeo que serán tráileres. Ambas categorías contendrán 3 estímulos diana correspondientes a los tres videojuegos de interés y otros 3 estímulos ruido. En la parte de valoración de cada muestra, el sujeto de evaluación realizará un pequeño test post-juego para analizar las bondades a nivel de experiencia del juego y de relaciones con los elementos.

PRIMERAS PRUEBAS

Con objeto de detectar fallos en la metodología y, muy especialmente, de ver la viabilidad del montaje experimental propuesto se han realizado unas primeras pruebas con usuarios. Las pruebas se llevaron a cabo con tres sujetos no ajenos al proyecto, en una única jornada intensiva en un laboratorio de *BitBrain Technologies* (ver Figura 4). La realización de estas primeras pruebas fue muy satisfactoria, ya que se logró poner en marcha todo el set-up tecnológico con éxito y se comprobó el correcto funcionamiento de todo el software de sincronización y toma de datos de los sensores (ver Figura 5). Para no alargar las pruebas se calibró y evaluó sólo en una dimensión, la de activación. Se consiguieron realizar las partes más importantes (Evaluación del Diseño y Evaluación de la Interacción) obteniendo datos de los sensores para su posterior procesamiento. Además, los sujetos voluntarios proporcionaron un valioso *feedback* que engloba aspectos de protocolo, de jugabilidad, y de carácter ergonómico a través de los comentarios realizados durante la evaluación y de un informe que redactó cada uno de ellos sobre la experiencia.



Figura 4. Vista del Laboratorio de pruebas durante la evaluación de la jugabilidad de las demos.

En los dos siguientes sub-apartados se recogen los problemas encontrados y las mejoras propuestas para solventarlos. Finalmente, se incluye un apartado que trata sobre los primeros análisis de los datos obtenidos durante las primeras pruebas.

Problemas encontrados

En las notas tomadas por los miembros del equipo evaluador y los informes de los usuarios se detectaron los siguientes puntos problemáticos:

- [P1] **Duración de las sesiones.** Si bien es cierto que en una primera estimación se ajustaron tiempos, durante el experimento se encontró con que la duración era excesiva para el sujeto de evaluación.
- [P2] **Sensores invasivos.** El *EyeTracker* usado, pese a ser una fuente de información potencialmente interesante, es una tecnología invasiva en cuanto a rasgos faciales se refiere. Se hicieron varias pruebas diferentes pero debido a su colocación sobre la cara y los reflejos producidos, en ninguna de ellas se consiguió que el seguimiento facial fuese el adecuado para la detección de la respuesta emocional facial del usuario (ver Figura 6). Por otra parte, el sensor de saturación de oxígeno ubicado en el dedo índice impedía al usuario sujetar el controlador correctamente.
- [P3] **Dificultad de autoevaluación emocional.** Los sujetos mencionan en sus informes la dificultad de evaluar su propio estado emocional, sin haber recibido algún tipo de entrenamiento con las escalas SAM. Además, la autoevaluación a lo largo del experimento depende en gran medida de la primera autoevaluación inicial realizada, normalmente llevada a cabo de manera imprecisa.
- [P4] **Transición entre estímulos.** Los sujetos expresan que resulta difícil pasar de un estímulo de sonido al siguiente (i.e. volver al estado neutro y prepararse para el siguiente estímulo).
- [P5] **Incomodidad.** En ocasiones, los sujetos se sentían abrumados al estar equipados con tanta tecnología y observadores en la sala.
- [P6] **Volumen de Datos.** Hay que tener en cuenta el despliegue tecnológico que supone una evaluación completa y la cantidad de datos que se están tomando durante la misma. En un determinado momento, hubo que volcar datos entre dispositivos de almacenamiento externo para no saturar los discos duros locales con toda la información que estaban recibiendo.



Figura 5. Sujeto de evaluación con la tecnología equipada

Mejoras necesarias

En este apartado se recogen aquellas recomendaciones y mejoras relacionadas directamente con los problemas encontrados en las primeras pruebas y descritos en el punto anterior. Estas deberán ser tenidas en cuenta para llevar a cabo la futura evaluación completa con usuarios.

- [S1] Reducción del número de estímulos tanto en calibración como en evaluación de categorías.
- [S2] Supresión del sensor *EyeTracker* (ET) por entrar en conflicto con el reconocedor facial y del sensor de saturación de oxígeno (SpO2).
- [S3] Sustitución de la escala SAM de autoevaluación emocional por una escala visual-analógica (EVA). Dicha escala es análoga al SAM, con la particularidad de que solo contiene indicaciones en los extremos de ésta, resultando menos confusa y más sencilla de utilizar.
- [S4] Ampliación del tiempo de descanso de las muestras para dejar al sujeto volver a un estado neutro emocional. En el caso de estímulos más largos como las demos, puede mostrarse al sujeto material audiovisual relajante.
- [S5] Presencia de un único evaluador en la sala que haga el papel de facilitador, siguiendo un protocolo de actuación perfectamente detallado, que aporte al usuario información constante sobre el proceso y el montaje experimental.
- [S6] Revisión de la necesidad de recopilar toda la información visual de forma continua; estudio de la reducción de la frecuencia de captura de las cámaras de vídeo y aplicación de técnicas de compresión de datos a los videos capturados.



Figura 6. Problemas de reflejos generados por el uso del *EyeTracker*.

Análisis inicial de los datos

Los datos obtenidos de las pruebas están siendo analizados con objeto de adecuar tanto los algoritmos de clasificación emocional facial [9] y fisiológica, como de fusión de los resultados de cada modalidad [8].

Los primeros análisis tanto de los datos faciales (ver Figura 7) como de los fisiológicos están siendo satisfactorios. Se ha podido comprobar la captura satisfactoria por los diferentes sensores de los diversos estados emocionales inducidos en los usuarios dispuestos durante la evaluación. Otro de los aspectos que cabe destacar es la reafirmación de la importancia de la utilización de sensores complementarios (tanto fisiológicos como no fisiológicos), que suplan los fallos en la clasificación emocional de una modalidad. Así, los sensores fisiológicos son capaces de recoger/detectar información emocional incluso cuando el usuario no la está mostrando explícitamente mediante sus expresiones

faciales. Y los faciales pueden ayudar a discernir entre estados emocionales que den lugar a respuestas fisiológicas similares.



Figura 7. Primeros resultados del analizador facial.
Usuario contento tras pasar uno de los niveles.

CONCLUSIONES Y TRABAJO FUTURO

En este artículo se ha presentado NeuroGame, un proyecto que recoge la experiencia de un equipo multidisciplinar especializado por un lado, en la evaluación hedónica de videojuegos y por otro, en la detección facial y fisiológica del estado emocional de los usuarios. El objetivo inicial del proyecto y en el que se ha centrado este artículo, es el desarrollo de una metodología para la medición y análisis de información asociada a la evaluación hedónica de videojuegos, abarcando no sólo la evaluación de la interacción, sino del diseño y comunicación del videojuego. Una vez realizada una primera prueba con tres usuarios, y comprobada la viabilidad del montaje experimental y de la metodología general propuestos, el siguiente paso será mejorar el protocolo existente para afrontar las futuras evaluaciones con usuarios. El objeto final del proyecto es contribuir al enriquecimiento de los métodos tradicionales de evaluación, haciéndolos presentes en las etapas más tempranas del desarrollo de videojuegos.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el Gobierno Español a través del DGICYT contrato TIN2011-24660.

REFERENCIAS

- Bradley, M.M. and Lang, P.J. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 49-59.
- Brave, S. and Nass, C. 2002. Emotion in human-computer interaction. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (chap. 4)*, 81-96. J. Jacko & A. Sears (Eds.) Hillsdale, NJ: Lawrence Erlbaum Associates.
- Costa, T., Rognoni, E. and Galati, D. 2006. EEG phase synchronization during emotional response to positive and negative film stimuli. *Neuroscience Letters* 406, 159-164.
- Cuthbert B.N., Schupp H.T., Bradley M.M., Birbaumer N. and Langley P.J. 2000. Brain potentials in affective picture processing covariation with autonomic arousal and affective report. *Biological Psychology* 52, 95-111.
- Ekman, P., Dalglish, T. and Power, M. 1999. *Handbook of Cognition and Emotion*. Wiley Online Library.
- González Sánchez, J.L., Gil Iranzo, R.M. and Gutiérrez Vela, F. L. 2011. Enriqueciendo la Evaluación en Videojuegos. *Actas del XII Congreso Internacional de Interacción Persona-*

- Ordenador. Interacción 2011* (Lisboa, Portugal, 2-6 de Septiembre). Néstor Garay y Julio Abascal (Eds.), 273-282.
- Gunes, H. and Pantic, M. 2010. Automatic, dimensional and continuous emotion recognition. *Int. Journal of Synthetic Emotions (IJSE)*, 1(1), 68-99.
- Hupont, I., Ballano, S., Cerezo, E. and Baldassarri, S. 2011. Scalable multimodal fusion for continuous affect sensing. In *Proceedings of IEEE Workshop on Affective Computational Intelligence - WACI* (Paris, 11-15 April), 1-8.
- Hupont, I., Cerezo, E. and Baldassarri, S. 2010. Sensing Facial Emotions in a Continuous 2D Affective Space. In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics – SMC 2010*, 2045-2051.
- Kim, Y.O., Paik, J.K., Heo, J., Koschan, A., Abidi, B. and Abidi, M. 2003. Automatic face region tracking for highly accurate face recognition in unconstrained environments. In *Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS'03)*, 29-36.
- Law, E.L-C., Roto, V., Hassenzahl, M., Vermeeren, A. and Kort, J. 2009. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of CHI 2009* (Boston, USA, April 04-09), 719-728.
- Liu Y., Sourina, O. and Nguyen, M. K. 2010. Real-time EEG-based Human Emotion Recognition and Visualization. In *Proc. 2010 Int. Conf. On Cyberworlds*. (Singapore, 20-22 Oct), 262-269.
- Maaoui, C. and Pruski, A. 2010. Emotion Recognition through Physiological Signals for Human-Machine Communication. *Cutting Edge Robotics (chap. 20)*, Vedran Kordic (Ed.) 317-332.
- Mandryk, R., Inkpen, K. and Calvert, T.W. 2006. *Using psychophysiological techniques to measure user experience with entertainment technologies*. Behaviour & Information Technology, 25(2), 141 – 158.
- Nasoz, F., Alvarez, K., Lisetti, C.L. and Finkelstein, N. 2003. Emotion recognition from physiological signals for presence technologies. *International Journal of Cognition, Technology and Work, Special Issue on Presence*, 6(1), 1-32.
- Norman, D. 2002. *Emotion and Design: Attractive things work better*. Interactions Magazine, 9(4), 36-42.
- Picard, R.W. 1997. *Affective Computing*. MIT Press
- Ravaja, N., Salminen, M., Holopainen, J., Saari, T., Laarni, J. and Järvinen, A. 2004. Emotional response patterns and sense of presence during video games: potential criterion variables for game design. In *Proceedings of the third Nordic conference on Human-Computer Interaction* (Tampere, Finland), 339-347.
- Russell, J.A. and Mehrabian, A. 1977. Evidence for a three-factor theory of emotions. *Journal of Research in Personality* 11(3), 273-294.
- Wagner, J., Kim J. and André, E. 2005. From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification. *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference*, 940-943

BIBLIOGRAFÍA Y DOCUMENTACIÓN

- [ARFF] Attribute-Relation File Format. Página de la Universidad de Waikato, <http://www.cs.waikato.ac.nz/ml/weka/arff.html> (último acceso 09/07/2012)
- [Ball11] S. Ballano. “Evaluación de un Sistema Multimodal de Reconocimiento de Emociones”. Universidad de Zaragoza, 2011. Tesis Fin de Máster.
- [BL94] Bradley, M.M., Lang, P.J. 1994. “Measuring emotion: The self-assessment manikin and the semantic differential”. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 49-59. 1994.
- [BN02] Brave, S., Nass, C. “Emotion in human-computer interaction”. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications (chap. 4) 81-96. J.Jacko & A. Sears (Eds.) Hillsdale, NJ: Lawrence Erlbaum Associates. 2002.
- [CHF06] Y. Hang, C. Hu, R. Feris, M. Turk. “Manifold based analysis of facial expression”. *J. Image & Vision Computing*, Vol. 24, nº 6, pp. 605-614, 2006.
- [CSG03] I. Cohen, N. Sebe, A. Garg, L.S. Chen, T.S. Huang. “Facial expression recognition from video sequences: temporal and static modelling”. *J. Computer Vision and Image Understanding*, Vol. 11, pp. 160–187. 2003.
- [Dar98] C. Darwin. *The expression of the emotions in man and animals (3rd Edition)*. New York: Oxford University Press.
- [DBW07] Y. Du, W. Bi, T. Wang, Y. Zhang, H. Ai, "Distributing expressional faces in 2-D emotional space" *In Proceedings of the 6th ACM International Conference on Image and Video Retrieval, 2007*, pp. 395-400. 2007.
- [DCSC07] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. McRorie, J.-C. Martin, L. Devillers, S. Abrilian, A. Batliner, N. Amir, K. Karpouzis. “The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data”. *In Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction (ACII '07)*, vol. 4738 of *Lecture Notes in Computer Science*, pp. 488–500, Lisbon, Portugal. September 2007.
- [EIDE] Eclipse Integrated Development Environment. <http://www.eclipse.org/> (ultimo acceso el 22/0/2012)
- [Ekm99] P. Ekman. *Facial Expression, The Handbook of Cognition and Emotion*. John Wiley et Sons, 1999.
- [ELop] Edouard López. Documento comparativo de funcionalidades entre los software Weka y R. Url: <http://edouard-lopez.com/fac/ICPS%20-%20S7/Data%20Mining/ComparingWekaAndR.pdf> (último acceso el 22/07/2012)
- [FASM] Software FaceAPI, SeeingMachines. Url: <http://www.seeingmachines.com/product/faceapi/> (último acceso 26/06/2012).

- [FT05] N. Fragopanagos, J.G. Taylor. "Emotion recognition in human computer interaction" *Neural Networks*, 18(4), pp. 389-405. 2005.
- [GGG11] González Sánchez, J.L., Gil Iranzo, R.M, Gutiérrez Vela, F. L. 2011. "Enriqueciendo la Evaluación en Videojuegos". *Actas del XII Congreso Internacional de Interacción Persona-Ordenador. Interacción 2011* (Lisboa, Portugal, 2-6 de Septiembre). Néstor Garay y Julio Abascal (Eds.), 273-282.
- [GMK06] G. Garidakis, L. Malatesta, L. Kessous, N. Amir, A. Paouzaïou, K. Karpouzis. "Modeling naturalistic affective states via facial and vocal expressions recognition". In *Proceedings of the 8th international conference on Multimodal interfaces (ICMI 2006)*. Banff, Alberta, Canada, pp. 146-154. 2006
- [GP06] H. Gunes, M. Piccardi. "A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior". In *International Conference on Pattern Recognition (ICPR)*, volume 1, pages 1148–1153. 2006.
- [GP09] H.Gunes , M.Piccardi. "Automatic temporal segment detection and affect recognition from face and body display". *IEEE Transactions on Systems, Man and Cybernetics – Part B* 39(1) pp 64-84. 2009.
- [GP10] H.Gunes, M.Pantic. "Automatic, dimensional and continuous emotion recognition". *Int Journal of Synthetic Emotions (IJSE)* 1(1), 68-99. 2010.
- [HBC11] I. Hupont, S. Ballano, E. Cerezo, S. Baldassarri. "Scalable multimodal fusion for continuous affect sensing". In *Proceedings of IEEE Workshop on Affective Computational Intelligence - WACI (Paris)* pp 1-8. April 2011.
- [HCC05] Z. Hammal, L. Couvreur, A. Caplier, M. Rombaut, "Facial Expressions Recognition Based on The Belief Theory: Comparison with Diferent Classifiers". In *Proc. 13th International Conference on Image Analysis and Processing (ICIAP)*, Italy. 2005.
- [HH97] C.L. Huang, Y.M. Huang, "Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification", *J. Visual Comm. and Image Representation*, 8(3), pp. 278-290. 1997.
- [Hup10] I. Hupont, 2010. "Affective computing: emotional facial sensing and multimodal fusion". Universidad de Zaragoza, 2010. Tesis Doctoral.
- [KCT00] T. Kanade, J. Cohn, Y. Tian. "Comprehensive database for facial expression analysis". In *Proceedings of International Conference on Face and Gesture Recognition*, pp 46–53. March, 2000.
- [LBA99] M.J. Lyons, J. Budynek, S. Akamatsu, "Automatic Classification of Single Facial Images", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(12), pp. 1357-1362. 1999.
- [LCK10] P. Lucey, J.F Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews. "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression". In *Computer vision and Pattern Recognition Workshop on Human-Communicative Behavior*, 2010.
- [LS11] J. Lichtenauer, M. Soleymani. "MAHNOB-HCI-TAGGING Database Manual". London, 2011.

- [MKin] Microsoft Kinect. Url: <http://www.microsoft.com/en-us/kinectforwindows/> (último acceso 26/06/2012).
- [MPEG4] ISO/IEC 14496 – MPEG-4 International Standard, Moving Picture Experts Group. Url: www.cselt.it/mpeg (último acceso 26/06/2012).
- [MVP10] G. McKeown, M. Valstar, M. Pantic, R. Cowie. "The SEMAINE corpus of emotionally coloured character interactions". In *Proc. IEEE Int. Conf. Multimedia & Expo* pp. 1–6. 2010.
- [Nor02] Norman, D. "Emotion and Design: Attractive things work better". *Interactions Magazine*, 9(4), pp 36-42. 2002.
- [OpCV] Librerías OpenCV. Url: <http://opencv.willowgarage.com/wiki/> (último acceso 09/07/2012).
- [Pic97] R. Picard, 1997. *Affective Computing*. MIT Press, Cambridge.
- [Plu84] Plutchik, R. "Emotions: A general psychoevolutionary theory". In K.Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 197-219). Hillsdale, NJ: Lawrence Erlbaum Associates. 1984
- [PP06] M. Pantic, I. Patras, "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences". *IEEE Trans. On Systems, Man and Cybernetics - Part B*, 36(2), pp. 433-449. 2006.
- [PR00] M. Pantic, L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: The State of the Art". *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22(12), pp. 1424–1445. 2000.
- [PVR05] M. Pantic, M. Valstar, R. Rademaker, L. Maat. "Fully automatic Facial Action Recognition in Spontaneous Behavior". In *Proceedings of the International Conference on Multimedia and Expo*, pages 317–321, 2005.
- [Rmin] Rapid Miner. Url: <http://rapid-i.com/content/view/281/225/lang,en/> (último acceso el 22/07/2012).
- [RPro] Software R. Url: <http://www.r-project.org/> (último acceso el 22/07/2012).
- [Rus80] Russell, J. A. "A circumplex model of affect". *Journal of Personality and Social Psychology*, 39, pp 1161-1178. 1980.
- [SBB02] T. Sim, S. Baker, M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database," In *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*. May 2002.
- [SJC96] M. C. Su, W. F. Jean, H. T. Chang., "A Static Hand Gesture Recognition System Using a Composite Neural Network." *Fifth IEEE Int. Conf. on Fuzzy Systems*, pp. 786-792, New Orleans, U.S.A. (NSC85-2213-E-032-009). 1996.
- [SLPP12] M.Soleymani, J.Lichtenauer, T.Pun, M.Pantic "A Multi-Modal Database for Affect recognition and Implicit Tagging," *IEEE Trans. Affective Computing*, 3(1), pp. 42-55. Jan.-Mar, 2012.
- [SSB09] N. Stoiber, R Seguier, G. Breton, "Automatic design of a control interface for a synthetic face" In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, pp. 207-216. 2009
- [SSJ01] Scherer, K. R., Schorr, A. Johnstone, T. *Appraisal processes in emotion: Theory, methods, research*. New York: Oxford University press. 2001.

- [Tom62] Tomkins, S. S. *Affect, imagery, consciousness: Vol 1. The positive affects*. New York: Springer, 1962.
- [VDUB] Software Virtual Dub. Url: <http://www.virtualdub.org/> (ultimo acceso el 25/07/2012).
- [WBTT] Página web de la empresa BBT (BitBrain Technologies). Url: <http://bitbrain.es/> (último acceso 19/07/2012).
- [WCV] Wikipedia: Validación Cruzada. [http://en.wikipedia.org/wiki/Cross-validation_\(statistics\)](http://en.wikipedia.org/wiki/Cross-validation_(statistics)) (último acceso 10/07/2012).
- [WDoc] Documentación de Weka. Url: http://www.cs.waikato.ac.nz/ml/weka/index_documentation.html (último acceso el 22/06/2012).
- [WDPS] Departamento de Psicología y Sociología de universidad de Zaragoza. Url: <http://psico-socio.unizar.es/> (último acceso 22/06/2012)
- [WEKA] Waikato Environment for Knowledge Analysis. Url: <http://www.cs.waikato.ac.nz/ml/weka/index.html> (último acceso 09/07/2012).
- [WGNU] GNU. Página de GNU en Wikipedia. Url: <http://es.wikipedia.org/wiki/GNU> (último acceso 22/06/2012).
- [WGri] Página web del grupo GRIHO (Grupo de Investigación en Interacción Persona Ordenador e Integración de Datos). Url: http://griho.udl.cat/es/inici.html?__locale=es (último acceso 19/07/2012).
- [Whi89] Whissell, C. M. "The dictionary of affect in language". In R. Plutchik & H. Kellerman (Ed.). *Emotion: Theory, research and experience. The measurement of emotions. Vol 4, pp 113-131*. New York: Academic Press. 1989.
- [WXML] Estándar XML del W3C: Url: <http://www.w3.org/XML/> (último acceso 19/07/2012).
- [ZJ05] Y. Zhang , Q. Ji, "Active and dynamic information fusion for facial expression understanding from image sequence". *IEEE Trans. Pattern Analysis & Machine Intelligence*, 27(5), pp. 699-714. 2005.

URLS DE BASES DE DATOS EMOCIONALES

- [BDE-01] Página web de la BDE MAHNOB-HCI Tagging Database. Url: <http://mahnob-db.eu/> (último acceso 22/06/2012).
- [BDE-02] Página web de la BDE Humaine Database. Url: <http://sspnet.eu/2009/12/humaine-database> (último acceso 22/06/2012).
- [BDE-03] Página web de la BDE Belfast Naturalistic Database. Url: <http://belfast-naturalistic-db.sspnet.eu> (último acceso 22/06/2012).
- [BDE-04] Página web de la BDE SAL Database. Url: <http://www.image.ntua.gr/ermis> (último acceso 22/06/2012).
- [BDE-05] Página web de la BDE Oreiestea Database. Url: <http://manolito.image.ece.ntua.gr/oresteia/index.html> (último acceso 22/06/2012).
- [BDE-06] Página web de la BDE XM2VTSDB Database. Url: <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb> (último acceso 22/06/2012).
- [BDE-07] Página web de la BDE The AR Face Database. Url: <http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html> (último acceso 22/06/2012).
- [BDE-08] Página web de la BDE CVL Face Database. Url: <http://www.lrv.fri.uni-lj.si/facedb.html> (último acceso 22/06/2012).
- [BDE-09] Página web de la BDE The Psychological Image Collection at Stirling Database. Url: <http://pics.psych.stir.ac.uk> (último acceso 22/06/2012).
- [BDE-10] Página web de la BDE The Japanese Femal Facial Expression (JAFFE) Database. Url: <http://www.kasrl.org/jaffe.html> (último acceso 22/06/2012).
- [BDE-11] Página web de la BDE CMUPIE (Pose, Illumination and Expression) Database. Url: http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261 (último acceso 22/06/2012).
- [BDE-12] Página web de la BDE Indian Institute of Technology Kanpur Database. Url: <http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase> (último acceso 22/06/2012).
- [BDE-13] Página web de la BDE The Yale Face Database. Url: <http://cvc.yale.edu/projects/yalefaces/yalefaces.html> (último acceso 22/06/2012).
- [BDE-14] Página web de la BDE Caltech Frontal Face Database. Url: <http://www.vision.caltech.edu/html-files/archive.html> (último acceso 22/06/2012).
- [BDE-15] Página web de la Human Scan BioID Face Database. Url: <https://www.bioid.com/download-center/software/bioid-face-database.html> (último acceso 22/06/2012).
- [BDE-16] Página web de la BDE Ourlu University Physics-Based Face Database. Url: <http://www.ee.oulu.fi/research/imag/color/pbfd.html> (último acceso 22/06/2012).
- [BDE-17] Página web de la BDE The Bosphorus Database. Url: <http://bosphorus.ee.boun.edu.tr/Home.aspx> (último acceso 22/06/2012).

- [BDE-18] Página web de la BDE The Database of Faces AT&T. Url:
<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> (último acceso 22/06/2012).

- [BDE-19] Página web de la BDE The Yale Face Database B. Url:
<http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html> (último acceso 22/06/2012).

- [BDE-20] Página web de la BDE FABO Database. Url: <http://research.it.uts.edu.au/cvrg/FABO.htm>
(último acceso 22/06/2012).

- [BDE-21] Página web de la BDE Semaine Database. Url: <http://semaine-db.eu> (último acceso 22/06/2012).

- [BDE-22] Página web de la BDE MMI Database. Url: <http://www.mmifacedb.com/accounts/register/>
(último acceso 22/06/2012).

- [BDE-23] Página web de la BDE MUG Database. Url: <http://mug.ee.auth.gr/fed/> (último acceso 22/06/2012).

- [BDE-24] Página web de la BDE Cohn Kanade Database. Url:
http://www.ri.cmu.edu/research_project_detail.html?project_id=421&menu_id=261 (último acceso 22/06/2012).

- [BDE-25] Página web de la BDE Extended Cohn Kanade Database. Url:
http://www.pitt.edu/~jeffcohn/CVPR2010_CK+2.pdf (último acceso 22/06/2012).

