Evaluación y Uso del Estado Emocional en Entornos Educativos Interactivos

VNIVERSITAT DE VALÈNCIA

Departamento de Informática

Programa de Doctorado:

Informática y Matemática Computacional

Marzo 2017

Doctorando: D. Luis Marco Giménez

Director: Dr. D. Miguel Arevalillo Herráez

D. **Miguel Arevalillo Herráez**, Profesor Titular de Universidad del Departament d'Informàtica de la Universitat de València,

CERTIFICA QUE

La presente Tesis Doctoral original de **D. Luis Marco Giménez** titulada "**Evaluación y uso del estado emocional en entornos educativos interactivos**", ha sido realizada bajo mi dirección y supervisión y, a mi juicio, reúne los requisitos para su lectura y obtención del grado de Doctor.

Y para que así conste a los efectos oportunos, firmo el presente certificado en Valencia, a 31 de mayo de 2017.

Dr. Miguel Arevalillo Herráez

Your intellect may be confused, but your emotions will never lie to you

Su intelecto podrá estar confundido, pero sus emociones nunca le mentirán

Roger Ebert

Agradecimientos

Quiero agradecer la inestimable dedicación de mi amigo y director de tesis, Miguel Arevalillo, porque sin su ayuda no hubiera sido posible realizar este trabajo.

A Adrián Pérez por ayudarme en el formato de edición del documento final.

A mi buen amigo Joan Benavent por haber sido fuente de inspiración.

A mis padres por haber confiado siempre en mí.

Por último, especialmente, quiero agradecer todo el apoyo, amor y fuerza moral que me ha dado mi esposa, Cristina, para que consiga acabar esta tesis.

Gracias a todos por hacer realidad este sueño.

Resumen

En la actualidad, la enseñanza tradicional está convergiendo progresivamente hacia entornos educativos interactivos semi-presenciales o a distancia, encontrándose una oferta cada vez mayor, de formación *on-line*. En este sentido, los sistemas de tutorización inteligente (STI) aportan un valor añadido a la educación interactiva. Su objetivo principal es proporcionar al alumno las ayudas necesarias para la consecución de los objetivos pedagógicos propuestos, tomando en consideración las capacidades intrínsecas del estudiante y el contexto particular en el que se desarrolla la actividad formativa.

Existen estudios donde se reportan sólidas evidencias de que no sólo las capacidades cognitivas del alumno y su contexto tienen un impacto en su rendimiento, sino que también el estado emocional del estudiante puede suponer un efecto significativo en su motivación y, por tanto, en su rendimiento. A pesar de estas evidencias, la mayoría de los STI no toman en consideración este hecho.

Asumiendo la conveniencia de incluir en los sistemas de aprendizaje interactivos métodos para la detección del estado afectivo del estudiante con el propósito de mejorar su aprendizaje, satisfacción y compromiso con los objetivos didácticos propuestos, en esta tesis se aborda el problema de la detección emocional desde diferentes perspectivas. Por un lado, mediante el empleo de técnicas de visión por computador sobre vídeos o imágenes estáticas. Por otro, a través del análisis de señales fisiológicas. Por último, con una aproximación de tipo s*ensor-free* fundamentada en el análisis de patrones de comportamiento de los estudiantes durante su interacción con un STI para el aprendizaje la aritmética y el álgebra lineal.

Para cada una de estas perspectivas, en esta tesis doctoral se han realizado diferentes aportaciones relativas a la detección emocional: 1) una extensión a una base de datos de vídeos anotados con información afectiva (FEEDB); 2) el desarrollo de una aproximación holística novedosa para el reconocimiento de expresiones faciales, denominada Eigenexpressions, fundamentada en la técnica de *Eigenfaces*; y 3) el análisis de la información que las señales procedentes de electroencefalografía pueden aportar en la detección emocional del usuario.

Desde el punto de vista de las aportaciones en la mejora instruccional, se diseñó un estudio experimental mediante el uso de STI para la enseñanza de la aritmética y el álgebra lineal, con el propósito de recabar información sobre el aprendizaje, el contexto formativo y el estado afectivo del estudiante durante su

interacción con el STI para, posteriormente, desarrollar un sistema de aprendizaje automático con capacidad para valorar el estado emocional del usuario sobre las dimensiones de la Dominancia, la Activación y la Valencia. De este modo, fue posible incorporar al STI capacidad para la evaluación del estado afectivo del estudiante con el objetivo de mantener al usuario en un estado que favoreciera su interés por el aprendizaje, determinando para ello el nivel de ayuda a proveer más adecuado en cada momento, con el fin último de mejorar su satisfacción y aprendizaje final.

Índice

Agradecimientos v

Resumen viii

Lista de Figuras xiii

Lista de Tablas xvii

Abreviaturas y Acrónimos xxii

I Fundamentos teóricos 1

1 Introducción 3
1.1. Motivación . . . 3
1.2. Objetivos . . . 5
1.3. Estructura de la tesis . . . 5

2 Estado del arte 9
2.1. Aplicaciones en educación . . . 10
2.2. Detección emocional . . . 13
2.2.1. Visión por computador . . . 13
2.2.2. Señales fisiológicas . . . 18
2.2.3. Patrones de comportamiento . . . 21
2.3. Conclusiones . . . 22

3 Aprendizaje automático 25
3.1. Clasificación y métodos . . . 26
3.1.1. Vecino más próximo . . . 26
3.1.2. Árboles de decisión . . . 27
3.1.3. Perceptrón multicapa . . . 29
3.1.4. Máquinas de soporte vectorial . . . 30
3.2. Regresión . . . 32
3.3. Reducción de la dimensionalidad sobre imágenes . . . 34
3.4. Análisis de series temporales: DTW . . . 36

3.5. Evaluación del rendimiento de los clasificadores 37
3.5.1. Validación cruzada 37
3.5.2. Medidas de evaluación 37
3.5.3. Matrices de confusión 38
3.5.4. Exactitud 38
3.5.5. Precisión 39
3.5.6. Cobertura 39
3.5.7. Valor-F 39
3.5.8. Curvas ROC 40

II Aportaciones en la detección emocional 41

4 Extensión de una base de datos de vídeos: FEEDB 43
4.1. Descripción de FEEDB 45
4.2. Extracción de datos 46
4.3. Aproximación propuesta para la detección de estados afectivos sobre FEEDB 53
4.4. Clasificación de las muestras y limitaciones 54
4.5. Extensión de FEEDB 59
4.6. Conclusiones 61

5 Detección emocional sobre imágenes estáticas: Eigenexpressions 65
5.1. Introducción 66
5.2. Aproximación propuesta: Eigenexpressions 67
5.2.1. Diseño de la fase de entrenamiento 67
5.2.2. Diseño de la fase de reconocimiento 68
5.3. Configuración de Eigenexpressions 69
5.3.1. Base de datos empleada 70
5.3.2. Configuración experimental 71
5.4. Evaluación de Eigenexpressions 72
5.5. Extensión de Eigenexpressions mediante máscaras de expresiones faciales 75
5.5.1. Formulación y creación de las máscaras 76
5.5.2. Entrenamiento mediante máscaras 77
5.5.3. Reconocimiento mediante máscaras 78
5.6. Evaluación de la extensión mediante máscaras 78
5.7. Conclusiones 81

6 Detección emocional sobre señales fisiológicas: Análisis de MAHNOB-HCI 85
6.1. Análisis de la base de datos MAHNOB-HCI 86
6.2. Selección y extracción de características EEG 88
6.3. Clasificación y análisis de resultados 91
6.4. Conclusiones 97

III Aportaciones en la mejora instruccional **99**

7 Incorporación de soporte afectivo a un STI **101**
7.1. Fundamentos de HBPS . . . 103
7.2. Diseño del núcleo afectivo del STI . . . 105
7.3. Construcción del núcleo afectivo del STI . . . 107
7.4. Implementación y análisis de los clasificadores afectivos . . . 115
7.5. Experimentación . . . 117
7.6. Análisis de resultados . . . 119
7.7. Conclusiones y Discusión . . . 121

IV Conclusiones y trabajos futuros **123**

8 Conclusiones y trabajos futuros **125**
8.1. Visión por computador . . . 126
8.2. Señales fisiológicas . . . 127
8.3. Patrones de comportamiento . . . 128
8.4. Detección emocional en entornos de *e-learning* . . . 129
8.5. Líneas de investigación futura . . . 130
8.6. Publicaciones resultantes . . . 131

Bibliografía **133**

Lista de Figuras

2.1. Componentes conceptuales habitualmente presentes en un STI tradicional 12
2.2. Rostro de Paul Ekman con indicaciones de los músculos faciales que intervienen en la composición de las AU1, 2, 4, 6 y 7 *(fuente www.paulekman.com)* 15

3.1. Diagrama genérico del proceso de clasificación 26
3.2. Ejemplo de la clasificación de una muestra mediante el algoritmo del vecino más próximo con $K = 3$ 27
3.3. Ejemplo de clasificación de un vehículo entre sus posibles categorías mediante un árbol de decisión 28
3.4. Estructura típica de un perceptrón 29
3.5. Estructura genérica de un perceptrón multicapa 30
3.6. Ejemplo del hiperplano óptimo definido para la separación de dos clases distribuidas en un espacio bidimensional 31
3.7. Ejemplos de regresión lineal 33
3.8. Ejemplo de regresion no lineal 33
3.9. Ejemplo de curvas ROC para tres diferentes sistemas de aprendizaje . 40

4.1. Ejemplo de una captura de un vídeo contenido en la primera versión de FEEDB y algunas de sus expresiones faciales 45
4.2. Algunos de los puntos en el plano 2D proporcionados por *Microsoft Face Tracking SDK* 47
4.3. Información angular de la posición de la cabeza para *pitch*, *roll* y *yaw (fuente msdn.microsoft.com)* 47
4.4. Modelo de AU detectadas por el *Face Tracking SDK* de Kinect y su equivalencia con el modelo Candide-3 *(fuente msdn.microsoft.com)* . . 48
4.5. Proceso de la extracción de datos de los vídeos XED de FEEDB en dos etapas 50
4.6. Herramienta gráfica usada en la visualización y conversión de los datos binarios obtenidos de las grabaciones originales en formato XED de FEEDB 51
4.7. Espacio de la cámara del sensor Kinect medido en metros y representación en un espacio de vídeo estándar de 640x480 píxeles de resolución 52

4.8. Visualización de los datos de FEEDB extraídos con Kinect en Mathworks MATLAB 53
4.9. Comparativa de las secuencias de movimientos de las AU para dos estados diferentes (placer y bostezo) y distintos sujetos 55
4.10. Comparativa de las secuencias individuales de movimientos de cada AU para dos estados diferentes (placer y bostezo) y distintos sujetos . 56
4.11. Comparativa de las secuencias de movimientos de las AU para un mismo estado (placer) y distintos sujetos 57
4.12. Comparativa de las secuencias individuales de movimientos de cada AU para un mismo estado (placer) y distintos sujetos 58

5.1. Fronteras de separación entre clases 68
5.2. Primera etapa de entrenamiento en Eigenexpressions 69
5.3. Ejemplo de *Eigenfaces* obtenidas y reducidas mediante PCA para cada una de las seis emociones básicas 70
5.4. Segunda etapa de entrenamiento en Eigenexpressions 71
5.5. Etapa de reconocimiento en Eigenexpressions 72
5.6. Comparativa de los valores de cobertura por clase entre *Eigenfaces* y Eigenexpressions para un tamaño de entrenamiento $w = 9$ 74
5.7. Comparativa de los valores de cobertura por clase entre *Eigenfaces* y Eigenexpressions para un tamaño de entrenamiento $w = 14$ 75
5.8. Comparativa de los valores de cobertura por clase entre *Eigenfaces* y Eigenexpressions para un tamaño de entrenamiento $w = 17$ 76
5.9. Ejemplo de las máscaras obtenidas con *Eigenfaces* para cada una de las seis emociones básicas 78
5.10. Comparativa de los valores de cobertura por clase entre *Eigenfaces*, Eigenexpressions y la aproximación basada en máscaras, para un tamaño de entrenamiento con $w = 9$ 80
5.11. Comparativa de los valores de cobertura por clase entre *Eigenfaces*, Eigenexpressions y la aproximación basada en máscaras, para un tamaño de entrenamiento con $w = 14$ 81
5.12. Comparativa de los valores de cobertura por clase entre *Eigenfaces*, Eigenexpressions y la aproximación basada en máscaras, para un tamaño de entrenamiento con $w = 17$ 82

6.1. Posicionamiento de los 32 electrodos en el casco BIOSEMI según el estándar internacional 10-20 89
6.2. Ejemplo de una muestra de las señales correspondientes a los 32 electrodos del EEG procesados con MATLAB y EEGLAB 90
6.3. Ejemplo de visualización de la actividad de cada uno de los 32 electrodos del EEG como valores del logaritmo de la potencia espectral y de la frecuencia 91
6.4. Representación de las 24 emociones recogidas en el proyecto GRID sobre las dimensiones de la Valencia y la Activación (*fuente Johnny Fontaine 2013*) 96

6.5. Distribución de las muestras EEG de MAHNOB-HCI cuantificadas con los valores del proyecto GRID y clasificadas mediante SVR 97

7.1. Interfaz gráfico de HBPS 104
7.2. Test SAM utilizado para capturar los estados afectivos de los estudiantes 108
7.3. Área ROC (0,694) obtenida para el clasificador que aproxima f'_{M_1} (Dominancia) 116
7.4. Área ROC (0,746) obtenida para el clasificador que aproxima f'_{M_2} (Valencia y Activación) 116

Lista de Tablas

2.1. Relación entre las seis emociones básicas y las AU relevantes que intervienen en la formación de sus correspondientes expresiones faciales según EMFACS 15

3.1. Ejemplo de una matriz de confusión para la clasificación de dos posibles clases 38
3.2. Distribución de aciertos y fallos para la clasificaicón de dos posibles clases 38
3.3. Ejemplo de una matriz de confusión con resultados sesgados hacia una clase 39

4.1. Resumen de características de las dos versiones de FEEDB 45
4.2. Interpretación de los valores de las AU reportadas por el *Face Tracking SDK* de Kinect 49
4.3. Enumeración y descripción de las SU detectadas por el *Face Tracking SDK* de Kinect y su equivalencia con las SU definidas en el modelo Candide-3 50
4.4. Lista de ficheros y elementos de información generados para cada grabación XED de FEEDB 53
4.5. Grabaciones y estados afectivos / expresiones sobre las que se ha extendido la información existente en FEEDB 61
4.6. Emociones y estados afectivos de FEEDB agrupados en categorías . . 63
4.7. Emociones y estados de FEEDB clasificados en clases de Valencia . . 63
4.8. Emociones y estados de FEEDB clasificados en clases de Activación . 63

5.1. Comparativa de la exactitud global (*accuracy*) entre diversos esquemas de clasificación en Eigenexpressions para diferentes tamaños de w . . . 73
5.2. Comparativa del área ROC entre diversos esquemas de clasificación en Eigenexpressions para diferentes tamaños de w 73
5.3. Comparativa de la exactitud global (*accuracy*) de *Eigenfaces* frente a Eigenexpressions en el reconocimiento de expresiones faciales para diferentes tamaños de w 73
5.4. Comparativa de la exactitud global (*accuracy*) del método de múltiples subespacios frente a Eigenexpresssions para diferentes tamaños de entrenamiento w 74

5.5. Comparativa de la exactitud global (*accuracy*) de la aproximación basada en máscaras frente al método Eigenexpressions para diferentes tamaños de entrenamiento w 79
5.6. Comparativa de la exactitud global (*accuracy*) de la aproximación basada en máscaras frente al método estándar de *Eigenfaces* para diferentes tamaños de entrenamiento w 79
5.7. Comparativa de la exactitud global (*accuracy*) de la aproximación basada en máscaras frente al método estándar de *Eigenfaces* mediante múltiples subespacios para diferentes tamaños de entrenamiento w . . 79

6.1. Emociones y estados recogidos en MAHNOB-HCI 88
6.2. Estructura de los datos fisiológicos contenidos en los ficheros BDF de MAHNOB-HCI 88
6.3. Resultados obtenidos mediante diferentes estrategias de clasificación para las 9 emociones recogidas en MAHNOB-HCI 92
6.4. Cobertura media obtenida (valor-F entre paréntesis) en la clasificación de las emociones Neutral, Alegría y Diversión para MAHNOB-HCI . . 92
6.5. Resultados en la clasificación de las 9 emociones recogidas en MAHNOB-HCI tras realizar una selección de características mediante un test ANOVA 93
6.6. Distribución de las 9 emociones de MAHNOB-HCI en tres clases de Activación 93
6.7. Distribución de las 9 emociones de MAHNOB-HCI en tres clases de Valencia 93
6.8. Resultados de la cobertura media en la clasificación de las 9 emociones de MAHNOB-HCI sobre las dimensiones de Activación y Valencia . . 94
6.9. Resultados de la cobertura media en la clasificación de las 9 emociones de MAHNOB-HCI sobre las dimensiones afectivas de Activación y Valencia tras realizar una selección de características mediante un test ANOVA 94
6.10. Valores medios obtenidos mediante análisis factorial con rotación VARIMAX en las dimensiones de Valencia y Activación y adaptación para las 9 emociones recogidas en MAHNOB-HCI 95

7.1. Variables que definen el contexto para un estudiante particular $e_q \in E$ en la resolución de un problema concreto $p_r \in P$ 109
7.2. Nivel de consecución de las estrategias descritas para un estudiante particular $e_q \in E$ en la resolución de un problema concreto $p_r \in P$ . . 110
7.3. Entradas del clasificador para la estrategia M_1 (Dominancia) durante la fase de entrenamiento 112
7.4. Entradas del clasificador para la estrategia M_2 (Valencia y Activación) durante la fase de entrenamiento 113
7.5. Entradas de los clasificadores durante la fase de predicción 114
7.6. Rendimiento obtenido por los diferentes métodos de clasificación evaluados para la construcción de los clasificadores afectivos de HBPS, en términos de cobertura media y área ROC 115

7.7. Optimización de los clasificadores afectivos: Dominancia (f'_{M_1}), Valencia y Activación (f'_{M_2}) 115
7.8. Matriz de confusión obtenida en la aproximación de f'_{M_1} (Dominancia) 117
7.9. Matriz de confusión obtenida en la aproximación de f'_{M_2} (Valencia y Activación) 117
7.10. Aplicación de las estrategias de maximización para cada grupo de estudiantes y conjunto de problemas 118
7.11. Variaciones reportadas en la Dominancia ($\Delta D = D_{e_q,p_r} - D_{e_q,p_{r-1}}$) para cada una de las estrategias empleadas durante la resolución de los problemas 119
7.12. Variaciones reportadas en la Valencia ($\Delta V = V_{e_q,p_r} - V_{e_q,p_{r-1}}$) y en la Activación ($\Delta A = A_{e_q,p_r} - A_{e_q,p_{r-1}}$) para cada una de las estrategias empleadas durante la resolución de los problemas 119

Abreviaturas y Acrónimos

Abreviatura	Descripción
AgCl	Cloruro de plata
ANOVA	Análisis de la varianza (*ANalysis Of VAriance*)
AU	Unidades de acción (*Action Units*)
BDF	Formato de datos Biosemi (*Biosemi Data Format*)
DTW	Alineamiento Temporal Dinámico (*Dynamic Time Warping*)
ECG	Electrocardiograma
EEG	Electroencefalograma
EMFACS	Sistema emocional de codificación de acciones faciales (*Emotional Facial Action Coding System*)
EMG	Electromiograma
EOG	Electrooculograma
FACS	Sistema de codificación de acciones faciales (*Facial Action Coding System*)
FACSAID	Diccionario para la interpretación afectiva de FACS (*FACS Affect Interpretation Dictionary*)
GSR	Respuesta galvánica de la piel (*Galvanic Skin Response*)
GUI	Interfaz gráfico de usuario (*Graphical User Interface*)
HCI	Interacción computador-máquina (*Human-Computer Interface*)
HBPS	Solucionador de Problemas Basado en Hipergrafos (*Hypergraph Based Problem Solver*)
HMM	Modelos ocultos de Markov (*Hidden Markov Models*)
HR	Ritmo cardíaco (*Heart Rate*)
HRV	Variabilidad del ritmo cardíaco (*Heart Rate Variability*)
Hz	Hercio (*Hertz*)
ID3	*Iterative Dichotomiser 3*
kHz	Kilohercio (*Kilo Hertz*)
KPCA	Análisis de componentes principales basados en Kernel (*Kernel Principal Components Analysis*)
LDA	Análisis discriminante lineal (*Linear Discriminant Analysis*)

Abreviatura	Descripción
LMT	Árboles de regresión logística (*Logistic Model Trees*)
MKS	Microsoft Kinect Studio
PAD	Pleasure-Arousal-Dominance
PCA	Análisis de componentes principales (*Principal Component Analysis*)
RBF	Función de base radial (*Radial Basis Function*)
RMSE	Raíz del error cuadrático medio (*Root Mean Square Error*)
ROC	Característica operativa del receptor (*Receiver Operating Characteristic*)
SAM	*Self-Assessment Manikin*
SDK	Kit de desarrollo software (*Software Development Kit*)
SEM	Error estándar de la media (*Standard Error of the Mean*)
SSD	Disco de estado sólido (*Solid State Disk*)
SSE	Suma de los cuadrados del error residual (*Sum of Squares Error*)
STI	Sistema de tutorización inteligente (*Intelligent Tutoring System*)
SU	Unidades de forma y dimensión de partes del rostro (*Shape Units*)
SVM	Máquina de soporte vectorial (*Support Vector Machine*)
SVR	Regresión de vector de soporte (*Support Vector Regression*)
TSS	Suma de los cuadrados totales (*Total Sum of Squares*)

Parte I

Fundamentos teóricos

Capítulo 1
Introducción

Resumen
En este capítulo se introduce la motivación que dio lugar a este trabajo de tesis y los objetivos perseguidos en el mismo. La parte final del capítulo describe la estructura global del documento.

Contenidos

1.1. Motivación 3
1.2. Objetivos 5
1.3. Estructura de la tesis 5

Los sistemas educativos interactivos y, en particular, los tutorizados de forma inteligente (STI), son concebidos como herramientas de apoyo a la enseñanza que permiten adaptarse a las necesidades específicas del estudiante en un dominio particular de conocimiento y en un contexto de aprendizaje determinado, mediante la provisión de ayudas específicas para la consecución de los objetivos pedagógicos propuestos. Estos sistemas se diseñan con la intención de simular el comportamiento de un profesor o tutor tradicional, ofreciendo de forma personalizada al alumno las pautas, recomendaciones y ayudas más adecuadas a su nivel de conocimiento y de aprendizaje y al contexto educativo en el que se encuentra incurso, creando y evaluando en todo momento un modelo del estudiante típicamente fundamentado en su nivel de conocimiento y en su forma de aprendizaje.

1.1. Motivación

De modo similar a un entorno presencial, donde el profesor o tutor dispone de la capacidad para valorar aspectos adicionales a los exclusivamente cognitivos que podrían influir en el aprendizaje, como el estado emocional del alumno, es factible

pensar que dotar a los sistemas educativos interactivos de estas capacidades podría suponer una mejora en su rendimiento. Estudios previos han reportado sólidas evidencias de que el estado emocional del estudiante puede tener un impacto significativo en su motivación y, en consecuencia, en el rendimiento de su aprendizaje [3].

Por este motivo, se consideró importante evaluar el impacto de las estrategias instruccionales implementadas en un STI sobre las variaciones del estado afectivo de sus estudiantes, especialmente en entornos educativos interactivos realistas, donde es esencial que los dispositivos necesarios para la captura de la información que pueda dilucidar el estado emotivo sean poco intrusivos. En este sentido, la información que puede resultar relevante para determinar el estado afectivo del estudiante puede ser de naturaleza física, fisiológica o de comportamiento.

Desde la perspectiva de los aspectos físicos que pueden tener una relación con el estado afectivo del estudiante, se encuentran la caracterización de las expresiones faciales y el uso de técnicas de visión por computador para su reconocimiento y análisis. Estas técnicas suelen consistir, entre otras, en la detección de ciertas regiones faciales (ojos, boca, pómulos, etc.) y su análisis geométrico, en el seguimiento de los ojos para determinar en qué puntos el usuario fija su atención o en el análisis de la cara de forma completa para la comparación con expresiones faciales prototípicas. En general, estos enfoques suelen ser poco intrusivos y económicos, aspectos que los hacen plenamente viables para su aplicación en entornos educativos interactivos.

Desde el punto de vista de los factores fisiológicos que pueden influir en el estado emocional del usuario, las técnicas habitualmente empleadas en su análisis suelen comprender mediciones de señales corporales mediante electroencefalogramas, electrocardiogramas, temperatura, conductividad de la piel, patrones de respiración, etc. El problema que estas señales presentan es que los dispositivos necesarios para su medición suelen ser altamente intrusivos, además de costosos y complejos de calibrar adecuadamente.

Por último, desde el punto de vista del comportamiento, la detección de determinados patrones de comportamiento también puede aportar valiosa información relacionada con el estado afectivo del usuario. Información como la tasa de aciertos o errores cometidos durante la resolución de una determinada tarea o problema, el número de ayudas solicitadas en cada paso para alcanzar la resolución a un problema determinado o la latencia entre respuestas, son parámetros que se pueden recabar de forma sencilla y transparente para el estudiante mediante aproximaciones de tipo *sensor-free*, tales como el uso de datos de bitácora (ficheros *log*) y cuestionarios [53, 41, 155, 74, 45, 141, 140].

Entre las ventajas que suponen las aproximaciones de tipo *sensor-free* para la detección automática del estado afectivo del estudiante frente a soluciones basadas en sensores físicos se encuentran: 1) menor coste económico que el que puede suponer una aproximación basada en sensores; 2) menor intrusividad; 3) en ocasiones, los sensores no proporcionan la precisión requerida o resulta compleja su correcta ubicación y calibración; 4) el uso de sensores físicos pueden distraer a los usuarios y afectar al rendimiento de sus tareas; 5) no siempre es posible el uso de sensores en el aula o en entornos reales durante largos períodos de tiempo.

1.2. Objetivos

El objetivo principal de esta tesis es la evaluación del estado emocional del usuario sobre diferentes dimensiones afectivas y su aplicación en entornos educativos interactivos. En particular, se propone la implementación mediante una aproximación de tipo *sensor-free* y técnicas de aprendizaje automático (*machine learning*), de un sistema para la evaluación del estado afectivo del estudiante y su integración en un STI específicamente diseñado para el aprendizaje de la aritmética y el álgebra lineal mediante la aplicación del método cartesiano.

Las dimensiones afectivas a considerar son aquellas que pueden ser fácilmente recogidas mediante la auto evaluación del estado emocional del usuario a través de formularios SAM (*Self-Assessment Manikin*) [27]. Estas dimensiones corresponden a la Valencia (actitud positiva o negativa), la Activación (estados de relajación o de excitación) y la Dominancia (sentimiento de autonomía o de dependencia en la realización de las actividades propuestas).

El propósito final perseguido en este trabajo es alcanzar un compromiso a través del cual se puedan regular las dimensiones afectivas anteriormente definidas mediante técnicas de aprendizaje automático, con el objetivo de proporcionar a los estudiantes un modo más efectivo para la superación de las dificultades con las que éstos suelen encontrarse durante el aprendizaje, consiguiendo que su implicación con el programa educativo y la satisfacción en la consecución de los objetivos pedagógicos sean mejorados frente a sistemas educativos interactivos que no toman en consideración el estado afectivo del estudiante.

1.3. Estructura de la tesis

La tesis se ha organizado en cuatro partes diferenciadas, formadas por un total de ocho capítulos.

En la parte I, denominada Fundamentos teóricos, se recogen los fundamentos que permiten la contextualización y comprensión de las aportaciones realizadas en este trabajo de investigación. En particular, este bloque consta de tres capítulos:

- El presente capítulo proporciona una visión global del conjunto de la tesis. En él se exponen la motivación y los objetivos generales que han conducido a la elaboración del trabajo de investigación en el ámbito de la evaluación del estado afectivo de un usuario y su aplicación en entornos educativos interactivos.

- En el capítulo 2 se realiza una revisión del estado actual de los principales métodos y técnicas empleados en la detección emocional de un usuario desde dos vertientes diferenciadas: por un lado desde el área de la detección emocional; por otro, desde el de sus aplicaciones en el ámbito educativo. En la primera parte del capítulo se analiza el impacto que las emociones pueden suponer en el aprendizaje desde el punto de vista de los sistemas adaptativos basados en la tutorización inteligente. En la segunda parte se examinan los trabajos más relevantes relacionados con la detección emocional, desde tres

áreas no excluyentes: a) visión por computador; b) señales fisiológicas (EEG, ECG, temperatura corporal, etc.); y c) patrones de comportamiento que pueden dilucidar el estado afectivo del usuario.

- En el capítulo 3 se describen las técnicas de aprendizaje automático habitualmente empleadas en cualquier problema de reconocimiento de patrones, detallando especialmente las utilizadas en las aportaciones realizadas en esta tesis doctoral.

La parte II, denominada Aportaciones en la detección emocional, contiene los trabajos y las contribuciones realizadas en esta tesis en lo relativo a la detección del estado afectivo del usuario. Esta parte consta de tres capítulos:

- El capítulo 4 describe el análisis, evaluación y extensión de la base de datos multimodal FEEDB (*Facial Expression and Emotion Database*) [175, 176], donde la principal aportación es la extensión realizada al corpus con información adicional extraída de los ficheros en formato XED contenidos en la base de datos.

- En el capítulo 5 se describe una aproximación original desarrollada como un método de detección del estado afectivo del usuario mediante técnicas holísticas basadas en la apariencia de la cara, fundamentadas en la técnica estándar de Eigenfaces, sobre un conjunto de imágenes contenidas en la base de datos Cohn-Kanade+ (*The Extended Cohn-Kanade Dataset*) [115].

- En el capítulo 6 se expone el análisis llevado a cabo sobre las señales procedentes de electroencefalografía contenidas en la base de datos multimodal MAHNOB-HCI [169], con la intención de explorar las posibilidades que este tipo de información puede aportar en la detección del estado emocional del usuario.

En la parte III, denominada Aportaciones en la mejora instruccional, se detallan las contribuciones realizadas mediante la evaluación del estado afectivo del usuario y su aplicación en sistemas de tutorización inteligente (STI). Esta parte se compone exclusivamente de un único capítulo:

- En el capítulo 7 se exponen los fundamentos de un STI para la resolución de problemas algebraicos y aritméticos denominado HBPS (acrónimo de solucionador de problemas basado en hipergrafos o *Hypergraph Based Problem Solver* en inglés), utilizado para la recopilación de los datos necesarios para la implementación, mediante técnicas de aprendizaje automático, de un núcleo de soporte afectivo que permita al STI regular el estado anímico y emocional del alumno mediante el análisis de su modelo de aprendizaje, la información del contexto pedagógico y la valoración de sus dimensiones afectivas sobre la Dominancia, la Valencia y la Activación.

Por último, la parte IV denominada Conclusiones y trabajos futuros consta de un único capítulo:

- En el capítulo 8 se exponen las conclusiones finales alcanzadas tras la realización del estudio presentado y se resume globalmente el trabajo de investigación llevado a cabo. Asimismo, considerando que en cualquier investigación realizada siempre quedan horizontes por explorar, se establecen las posibles líneas de investigación futuras que podrían dar continuidad a este trabajo de tesis.

Capítulo 2

Estado del arte

Resumen

En este capítulo se realiza una revisión de la literatura relacionada con las principales técnicas propuestas en el campo de la computación afectiva para el reconocimiento automático de emociones mediante computador y se analiza su empleo en el ámbito de las aplicaciones educativas. En la primera parte se revisa, en el ámbito de la enseñanza, cómo el estado afectivo del estudiante puede afectar a su rendimiento, examinando algunas de las técnicas comúnmente empleadas en este contexto. En la segunda parte se analizan los métodos habituales utilizados en la detección del estado afectivo del usuario que podrían resultar de aplicación en diversos ámbitos.

Contenidos

2.1. Aplicaciones en educación 10
2.2. Detección emocional 13
2.3. Conclusiones 22

Desde que el término *Affective Computing* fuera introducido en el campo de la computación por Rosalind Picard en su estudio sobre sistemas capaces de reconocer, interpretar, procesar y simular emociones humanas [143], el creciente interés del reconocimiento automático del estado afectivo humano ha supuesto en la última década un auge de la computación afectiva.

Posiblemente, uno de los ámbitos con más interés en el reconocimiento del estado afectivo y emocional de un usuario se encuentre en la interacción computador-humano (HCI), con el desarrollo de sistemas "empáticos" que puedan ser capaces de responder adecuadamente en función del estado del usuario, creando un entorno de interacción más amigable y, en definitiva, haciendo que las tareas sean más agradables y productivas. Un amplio abanico de entornos, como los relativos al

marketing, la salud o la educación, entre muchos otros, podrían verse beneficiados por sistemas que incorporen capacidades de reconocimiento afectivo.

No obstante, aunque en los últimos años los progresos en este área han sido significativos, el reto de la detección emocional humana y su respuesta sigue siendo una tarea compleja, no exenta de imprecisiones y limitada cuando se compara con la percepción emocional realizada por nuestro principal rival en este área: el cerebro humano. Esta dificultad es inherente a la propia naturaleza de las emociones, frágil y elusiva, donde además diferentes estados emotivos pueden compartir fronteras de definición, ser dependientes del contexto e incluso existir notables variaciones entre individuos.

En este capítulo se abordará la detección afectiva y emocional de un usuario desde dos perspectivas diferenciadas. Por un lado, desde el ámbito educativo se examinará el impacto que las emociones y sentimientos de un estudiante pueden suponer en su rendimiento y aprendizaje; por otro lado, se examinarán y discutirán las técnicas habitualmente empleadas en la detección del estado afectivo y emocional del usuario.

2.1. Aplicaciones en educación

Aunque no existe una teoría fundamentada sobre cómo las emociones influyen en el aprendizaje, es razonable pensar que el rendimiento de un estudiante, y por consiguiente su aprendizaje, puede verse afectado por sus emociones y sentimientos [29].

En [83] se exploran las razones por las que un estudiante de secundaria o bachillerato es capaz de resolver un problema de matemáticas, por ejemplo. Entre ellas se encuentra la autorrecompensa por haber alcanzado la solución, la posibilidad de obtener una buena calificación, evitar un castigo o, simplemente, complacer a sus padres o a su profesor.

De modo similar, aunque desde la perspectiva de las emociones y sentimientos negativos, cuando un estudiante percibe que no es capaz de resolver correctamente un ejercicio o un examen, éste tiende a cuestionar su capacidad o incluso a sentirse inútil para los objetivos que se le plantean [100]. Es decir, afloran estados emocionales y mentales con una connotación negativa que puede influir del mismo modo en su continuidad y motivación con el aprendizaje.

En [84] se demuestra que con tan sólo unos leves cambios positivos en el estado anímico de una persona se generan pensamientos más creativos, flexibles, detallados y eficientes en la resolución de un problema.

Por todo ello, independientemente del motivo perseguido por un alumno que resuelve satisfactoriamente un problema o del alumno que ve mermada su incapacidad ante la imposibilidad de resolver una prueba, es evidente que existe un componente emocional que relaciona sensaciones (placenteras o negativas) con su rendimiento y motivación y, por tanto, con su aprendizaje. Es por esta relación por la que diferentes escenarios educativos consideran la integración de la dimensión cognitiva del estudiante con su dimensión afectiva, con el propósito de alcanzar

un aprendizaje adaptativo y personalizado como respuesta combinada a ambas dimensiones.

No obstante, la detección del estado emocional en entornos educativos es una tarea no exenta de complejidad principalmente por dos razones [160]: 1) el estado afectivo del alumno no sufre grandes cambios durante su aprendizaje; 2) los cambios producidos suelen presentar una intensidad menor que en otros contextos. Estos retos, inherentes a los escenarios educativos, suponen un *handicap* en la identificación de estados afectivos relevantes [124]. Adicionalmente, este tipo de aplicaciones se encuentra limitado a aquellas situaciones en las que el uso de sensores es factible, lo que ha motivado recientes aproximaciones de tipo *sensor-free*, tales como el uso de datos de bitácora (típicamente ficheros *log*) o cuestionarios [53, 41, 155, 74, 141, 140].

A pesar de las dificultades asociadas a los entornos educativos, existe un creciente interés en el desarrollo de sistemas de aprendizaje que tomen en consideración el estado afectivo del estudiante. Una revisión reciente de la literatura sobre sistemas educativos de carácter afectivo basados en *e-learning*, donde se discuten los objetivos perseguidos en cada uno de los 26 trabajos analizados, se realiza en [160].

En este sentido, de modo similar a las estrategias seguidas por un profesor en clase donde adapta su dinámica formativa en función del estado del auditorio para minimizar situaciones de aburrimiento, frustración o abandono, y más específicamente en el caso de la tutorización individualizada (alumno-profesor) donde la adaptación se realiza completamente personalizada al perfil cognitivo del alumno y a su estado afectivo o emocional [107], los sistemas de tutorización inteligente que incorporan soporte afectivo imitan el comportamiento del tutor analizando no sólo el perfil cognitivo del estudiante tutorizado, sino también su estado afectivo para intentar maximizar su experiencia en el aprendizaje a través de la aplicación de diferentes estrategias instruccionales como, por ejemplo, la adaptación del nivel de dificultad del problema, la provisión de ayudas, de mensajes explicativos o motivacionales, etc.

Los STI son programas basados en computador desarrollados para la enseñanza de contenidos relativos a áreas tan diversas como lo es el aprendizaje de la aritmética o el álgebra, los idiomas o la física, con el propósito de proporcionar las ayudas instruccionales necesarias para que el alumno pueda conseguir los objetivos pedagógicos propuestos, creando en su interacción con el estudiante un modelo que refleje la forma en que éste aprende, sus dificultades y habilidades. Este modelado del alumno, junto con un modelo que refleje el dominio de conocimiento objetivo del aprendizaje, un modelo didáctico que defina las estrategias instruccionales a emplear y un interfaz de usuario con el que estudiante y sistema interactuarán mutuamente, es lo que habitualmente define a un STI tradicional. En la figura 2.1 se muestra una representación de los componentes que lo conforman. Sin embargo, no todos los sistemas de tutorización inteligente presentan estos cuatro bloques o componentes conceptuales. Un sistema considerado como STI debe disponer, como mínimo, de un modelo del alumno para ser usado con el fin de adaptar las instrucciones proporcionadas al estudiante que le permitan alcanzar los objetivos pedagógicos propuestos [171].

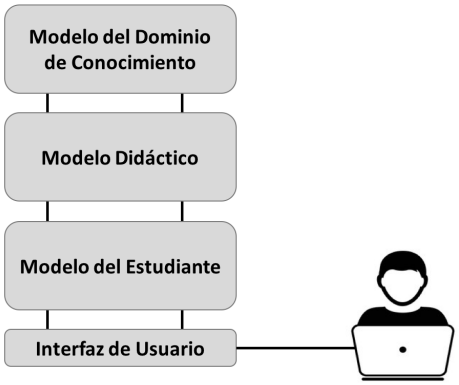


Figura 2.1: Componentes conceptuales habitualmente presentes en un STI tradicional

Es en el modelado del alumno donde, además de definir su conocimiento y su modo de aprendizaje, es factible recoger información de carácter afectivo como pueden ser sus sensaciones, percepciones y reacciones, de modo que el STI pueda disponer de información procedente de otras dimensiones diferentes a las propiamente cognitivas para adaptar su respuesta de forma individualizada al alumno.

En el diseño de cualquier STI que incorpore soporte afectivo es fundamental definir claramente qué se entiende como estado afectivo, cómo debe ser modelado internamente y cómo responder adecuadamente según el estado del estudiante. Es decir, no es lo mismo que el STI intente determinar un conjunto de estados concretos como la frustración, el aburrimiento o la felicidad, o que se limite a establecer valores para las dimensiones de Valencia, Activación y Dominancia. Las estrategias instruccionales del STI como respuesta a las acciones y el estado del alumno podrán tener mayor o menor granularidad dependiendo de los estados afectivos definidos por el STI.

En el mismo nivel de importancia se encuentra cómo el STI recaba la información afectiva del usuario. Su estado puede ser inferido mediante el uso de diferentes técnicas: análisis de la expresión facial, seguimiento de ojos, tamaño de las pupilas [92, 26, 188], movimientos corporales [131, 92], señales fisiológicas [51], patrones de comportamiento [45, 93] o autoevaluaciones procedentes de los usuarios [161, 51]. Cualquiera de ellas o la combinación de varias, podrá servir para que el STI disponga de información para predecir el estado afectivo del estudiante en cada momento. No obstante, todas las técnicas presentan sus limitaciones: las autoevaluaciones podrían no reflejar el verdadero estado emocional del estudiante por miedo a no cumplir las expectativas, mientras que las señales fisiológicas pueden presentar grandes variaciones entre diferentes personas o incluso ser malinterpretadas [172]; las técnicas basadas en visión pueden ser imprecisas, verse afectadas por la iluminación, por el fondo, por rotaciones de la cabeza u oclusiones; los patrones de comportamiento podrían no ser suficientes por sí mismos para la

determinación del estado emocional del estudiante.

2.2. Detección emocional

Diferentes aproximaciones se han propuesto para la detección del estado emocional por computador, como el reconocimiento de expresiones faciales mediante técnicas de visión artificial sobre imágenes estáticas o vídeos [89, 135, 102, 52, 181, 58, 30, 120], identificación de emociones a través del análisis del tamaño de la pupila [28, 104, 11], del análisis de la voz [154, 20, 58, 190], del lenguaje corporal [131, 35, 50], de señales fisiológicas mediante el análisis de electroencefalogramas (EEG) [5, 114, 169] u otras señales del sistema nervioso periférico como los electrocardiogramas (ECG), la temperatura corporal, la conductividad de la piel (GSR), los patrones sobre la respiración o la variabilidad del ritmo cardíaco (HRV) [95, 186, 169, 177].

Los sistemas que intentan detectar el estado afectivo de un usuario basándose en una única fuente de información, como la imagen o el audio, son considerados sistemas unimodales. Sin embargo, en un entorno cotidiano las personas expresan y comunican sus emociones y estados afectivos al resto de interlocutores a través de varios canales simultáneamente, como por ejemplo mediante información procedente de la cara (expresiones faciales), del habla tanto con información explícita o lingüística (el mensaje) como implícita o paralingüística (características prosódicas como el tono de la voz, la intensidad, la velocidad o el ritmo) y del cuerpo (gestos de las manos y posturas o movimientos del cuerpo). Los sistemas que consideran varias fuentes sincronizadas de información para la determinación del estado afectivo del usuario, combinando varias de las estrategias anteriormente citadas, son denominados sistemas afectivos multimodales. Algunos ejemplos de ellos se describen en [136, 165, 76, 12, 58, 54, 169].

2.2.1. Visión por computador

Dado que el rostro es el principal componente expresivo de los seres humanos, es comprensible que gran parte de los sistemas de detección emocional concentren sus esfuerzos en el análisis de la cara y de sus expresiones faciales [163].

Uno de los grandes problemas que presentan la mayoría de los sistemas de detección emocional mediante visión por computador es que básicamente trabajan sobre emociones prototípicas, exageradas y no espontáneas [115, 180], circunstancia que no se adecua a entornos de interacción realistas entre personas. Otros retos habituales con los que tiene que tratar el análisis de expresiones faciales mediante técnicas de visión y que pueden tener un gran impacto en el reconocimiento final son: 1) las variaciones lumínicas a las que puede estar expuesto el sujeto y que pueden arrojar sombras en ciertas partes del rostro que hacen compleja la detección de sus rasgos faciales. En este caso, las condiciones lumínicas de un laboratorio o estudio se encuentran habitualmente controladas, a diferencia de lo que suele suceder con luz natural donde el ángulo de iluminación puede afectar dramáticamente a la imagen (y de modo similar el fondo de la imagen capturada); 2) las variaciones en

la posición de la cabeza o movimientos espontáneos asociados a ciertas emociones pueden dificultar el reconocimiento de los componentes individuales de la cara; 3) las oclusiones debidas a giros de la cabeza, a accesorios como gafas o pañuelos o, incluso, a otras partes del cuerpo como puede ser una mano cubriendo total o parcialmente la boca o parte del cabello sobre el rostro (pelo o barba); 4) las diferencias individuales entre sujetos como la forma de la cara, el color o la textura de la piel, rasgos étnicos o la edad.

Sin embargo, una de las grandes ventajas de las técnicas de reconocimiento de expresiones faciales basadas en visión es que son muy poco intrusivas al emplear cámaras que no precisan de una fuerte interacción con el usuario. Además, son dispositivos razonablemente económicos. De hecho, la mayoría de las computadoras actuales disponen de una pequeña cámara incorporada en el monitor con suficiente resolución y calidad, que puede servir como herramienta para la adquisición y monitorización de los movimientos faciales del usuario mientras realiza otras tareas. Todos estos motivos han contribuido a que numerosas investigaciones hayan dirigido sus esfuerzos hacia la detección de las expresiones faciales y posterior determinación de las emociones mediante técnicas de visión.

Muchos de los sistemas desarrollados se centran en la detección de las seis emociones consideradas básicas [62]: Alegría, Asco, Ira, Miedo, Sorpresa y Tristeza, al ser consideradas éstas innatas, universales y no como un producto social de aprendizaje cultural. Estas emociones pueden ser descritas mediante la expresión facial representada por el sujeto en un período de tiempo normalmente breve, por lo que el reconocimiento de la expresión puede contribuir en la predicción de la emoción experimentada.

No obstante, en situaciones cotidianas reales estas seis emociones prototípicas no suelen aparecer de forma natural y habitual, ni con tanta frecuencia como otros estados afectivos como puede ser la fatiga, o estados mentales como la concentración, el aburrimiento, la confusión o la frustración, entre otros. Razones como éstas hacen que cada vez surjan más trabajos basados en el reconocimiento de emociones más allá de las seis emociones básicas [75, 91, 64, 92, 191, 87].

Las expresiones faciales mostradas como respuesta a cada emoción particular consisten en una combinación de movimientos de los músculos faciales configurados de una forma determinada para cada expresión. En esta línea, Ekman y Friesen desarrollaron en 1978 un sistema de codificación facial para medir los movimientos visibles de los músculos de la cara en términos anatómicos denominado FACS (*Facial Action Coding System*) [60, 78], basado en la definición de 44 unidades de acción independientes denominados AU o, en inglés, *Action Units*, a través de las cuales podían codificar los movimientos de los diferentes músculos faciales y, en consecuencia, caracterizar cualquier expresión facial que una persona puede realizar mediante la descomposición de la expresión en un conjunto de AU bien definido.

Mediante este sistema de codificación facial, Ekman y Friesen realizaron una categorización taxonómica de los movimientos faciales en función de las AU intervinientes en cada uno de ellos, así como las combinaciones de éstas. La mayoría de las AU descritas en FACS se relaciona con un solo músculo facial, como, por ejemplo, el caso de las AU1–AU4 que corresponden a los músculos que

intervienen en el movimiento de las cejas. En la figura 2.2 se muestra, a modo ilustrativo, el rostro de Paul Ekman con indicaciones de los músculos faciales que intervienen en la composición de las AU1, 2, 4, 6 y 7.

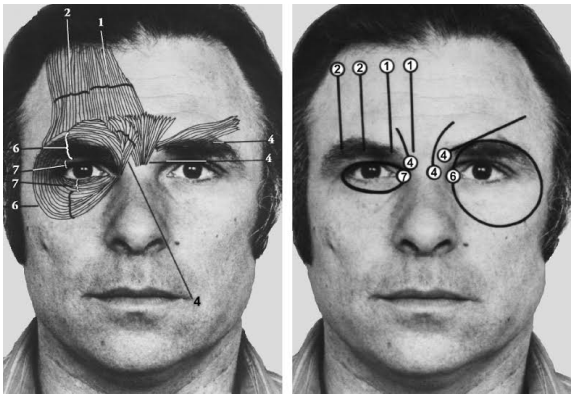


Figura 2.2: Rostro de Paul Ekman con indicaciones de los músculos faciales que intervienen en la composición de las AU1, 2, 4, 6 y 7 *(fuente www.paulekman.com)*

FACS es meramente descriptivo, es decir describe con precisión las expresiones faciales en términos de los movimientos visibles de las AU implicadas, pero no las relaciona con sus correspondientes emociones. Las relaciones entre expresiones y emociones se describen en sistemas como EMFACS (*Emotional Facial Action Coding System*) [72] o FACSAID (*Facial Action Coding System Affect Interpretation Dictionary*) [61]. En la tabla 2.1 se muestran las seis emociones básicas y la combinación de las AU más relevantes que intervienen en la formación de sus correspondientes expresiones faciales según se describe en EMFACS. En esta tabla, la emoción Felicidad se expresa como la deformación de la AU12 (AU12: estiramiento de los extremos de los labios o "*lip corner puller*" según el nombre descriptivo dado en FACS); la combinación de las AU6 y AU12 (AU6: elevación de las mejillas o "*cheek raiser*"); o la combinación de las AU7 y AU12 (AU7: fruncimiento de los párpados o "*Lid Tightener*").

Emoción	**AU relevantes en la formación de la expresión**		
Felicidad	12	6+12	7+12
Tristeza	1	1+4	
Sorpresa	1+2+5B	1+2+26	1+2+5B+26
Miedo	1+2+4	20	
Ira	4+5		
Asco	9	10	

Tabla 2.1: Relación entre las seis emociones básicas y las AU relevantes que intervienen en la formación de sus correspondientes expresiones faciales según EMFACS

Tanto FACS como los esquemas que relacionan expresiones y emociones son

los métodos más utilizados en el reconocimiento de expresiones faciales mediante sistemas de visión por computador [110, 56, 109, 138, 18].

Aunque una amplia variedad de los trabajos se han centrado en el análisis de expresiones exageradas y prototípicas [115, 180], debido principalmente a que las primeras bases de datos anotadas que recopilaban un número suficiente de imágenes o vídeos sobre expresiones faciales eran mayoritariamente posadas como en Cohn-Kanade [90], JAFFE (*The Japanese Female Facial Expression*) [117] y GENEVA [14] o con una combinación de imágenes y vídeos posados y espontáneos como en MMI [139], otros muchos estudios han dirigido sus líneas de investigación hacia el reconocimiento de expresiones faciales espontáneas. En este sentido, Zen et al. [192] realizan una amplia revisión de los métodos para el reconocimiento de expresiones espontáneas, aunque cada vez surgen más estudios que orientan sus líneas de investigación en esta dirección [19, 40, 16, 17, 183, 116, 166, 57, 126, 169, 97].

Un aspecto importante que debe considerarse en cualquier sistema que pretenda analizar el estado afectivo de un usuario, es que el espacio de la cara sobre el que analizar las expresiones faciales suele tener una dimensionalidad muy alta. Por este motivo, es habitual en los algoritmos de detección emocional procesen previamente las imágenes de las caras mediante técnicas lineales de reducción de la dimensionalidad como PCA (*Principal Component Analysis*) y LDA (*Linear Discriminant Analysis*) o de reducción no lineal como KPCA (*Kernel Principal Components Analysis*). Estás técnicas reducen la dimensionalidad de la imagen capturando la variabilidad de los datos, transformándolos a un subespacio de menor dimensionalidad que el inicial, disminuyendo de esta forma su almacenamiento en disco y memoria, el tiempo de procesado en los algoritmos de clasificación y aprendizaje automático, así como su eficiencia al reducir el impacto de la maldición de la dimensionalidad [82].

En lo relativo a los métodos habitualmente utilizados para la extracción y procesado de las características faciales como etapa previa al reconocimiento de la expresión facial y estimación de la emoción, éstos se pueden categorizar principalmente en dos grupos [180]

1. Modelos basados en características geométricas, en los que los diferentes componentes de la cara, como los ojos, las cejas, la nariz o la boca, son representados mediante vectores de parámetros geométricos con la definición de sus coordenadas, longitudes y ángulos, definiendo de esta forma las formas geométricas que los representan y, en consecuencia, la geometría de la cara [36, 137, 183, 102].

2. Modelos basados en la apariencia, donde se da más importancia a los valores de los píxeles que a la distancia relativa o forma entre componentes faciales. Parámetros como la intensidad de los píxeles que componen la cara o el histograma de la imagen son considerados por estos modelos, donde pueden ser aplicables diversos filtros sobre la imagen, como por ejemplo los *wavelets* de Gabor [19, 16, 17] o métodos basados en la apariencia activa (AAM) [59, 37, 116, 46].

A su vez, estos dos grupos se pueden dividir atendiendo a su ámbito de actuación en: 1) holísticos, los cuales consideran la imagen en su totalidad, en este caso la cara del individuo de forma completa; 2) locales, en los que se utilizan características o áreas concretas del rostro en las que se suelen producir deformaciones como consecuencia de una expresión facial, como la apertura de la boca, la contracción de los pómulos, la elevación de las mejillas, el arqueo de las cejas, la caída de la mandíbula, la apertura de los ojos o el estiramiento de las comisuras de los labios, entre otras.

Aproximaciones híbridas entre los dos modelos anteriormente descritos, es decir entre aquellos que usan características geométricas junto a modelos basados en la apariencia, pueden constituir alternativas con buenos resultados en el reconocimiento de expresiones faciales. En esta línea, diversos estudios han reportado mejoras en la combinación de modelos y técnicas [56, 179, 180]. Por otro lado, una extensa revisión actualizada y un análisis de los principales sistemas afectivos multimodales implementados hasta la fecha ha sido realizada por D'Mello y Kory [55]. Si bien esta revisión no se circunscribe en exclusiva a sistemas basados en el análisis de expresiones faciales, en ella se asevera que aunque los sistemas multimodales suelen obtener un rendimiento mayor que sus equivalentes unimodales en el reconocimiento de estados afectivos no espontáneos, estas mejoras no resultan tan evidentes cuando el análisis se realiza sobre expresiones espontáneas donde las mejoras son simplemente "modestas".

Aunque no son exclusivos de los sistemas afectivos basados en técnicas de visión artificial, sino que son aplicables a cualquier tipo de sistema que tenga que tomar una decisión, básicamente existen tres métodos para la fusión de información proveniente de diferentes fuentes, dependiendo del momento en el que se combine la información: 1) Fusión de datos, cuando la información de cada fuente es combinada directamente (*raw data*). Este tipo de fusión únicamente es aplicable a señales con la misma resolución temporal y su uso no es frecuente por los posibles problemas de sincronización que puede surgir entre dispositivos; 2) Fusión de características. En este caso la combinación se realiza sobre el conjunto de características obtenidas para cada señal de forma independiente. Su uso se encuentra más extendido por su simplicidad; 3) Fusión en el clasificador. Con este método las diferentes salidas de los clasificadores de cada señal individual se combinan para obtener la decisión final. Este método de fusión es el más utilizado en sistemas afectivos multimodales.

La mayoría de los modelos de reconocimiento de expresiones faciales han concentrado sus esfuerzos en la extracción de características en un espacio de representación bidimensional [15, 111, 109, 33], donde el análisis de la cara se realiza desde una perspectiva frontal o con poca variación en cuanto a rotación de la misma, mayoritariamente debido a que estos modelos 2D suelen fallar cuando el rostro no se encuentra frontalmente orientado.

La limitación de los modelos bidimensionales para la representación de la cara y el análisis de expresiones faciales es evidente, principalmente porque en entornos reales los cambios de expresión suelen venir también acompañados de movimientos y rotaciones en la posición de la cabeza. En este sentido, los modelos basados en una representación de la cara en 3D pueden resultar de utilidad especialmente en

sistemas de análisis de expresiones espontáneas donde, además de proveer mayor robustez frente a variaciones en la iluminación, suelen ser invariables a cambios en la pose [39, 187, 38, 166, 178].

Por último, es importante destacar el efecto de las micro-expresiones en el reconocimiento del estado afectivo del usuario [63], es decir de expresiones faciales de cortísima duración en el tiempo, habitualmente en el rango 1/12 – 1/20 segundos). Estas micro-expresiones, a diferencia de las faciales comunes, son difícilmente inhibidas de forma voluntaria por el sujeto, pudiendo revelar emociones ocultas. A este respecto, pocos estudios han analizado sus efectos [168, 167, 144, 189], principalmente debido a que son expresiones no verbales difícilmente visibles a simple vista incluso para un observador humano, unido a la falta de bases de datos especializadas en micro-expresiones, aunque en este sentido se ha elaborado recientemente SMIC (*Spontaneous Micro-expression Database*) [108] como base de datos anotada de micro-expresiones espontáneas que puede resultar de gran utilidad en la promoción del análisis de este tipo de expresiones faciales.

2.2.2. Señales fisiológicas

Es razonable pensar que una experiencia emocional puede suponer un cambio en el estado del cuerpo y, en consecuencia, de sus variables fisiológicas. Es fácil comprobar cómo una emoción como el Miedo puede afectar al cuerpo produciendo en algunos casos temblores musculares, así como un estado de nerviosismo puede afectar al habla o a la sudoración corporal. En [86] se asevera que los cambios producidos en el sistema fisiológico se encuentran estrechamente relacionados con la experiencia emocional percibida por el sujeto. Es por ello por lo que el análisis de estos cambios puede contribuir en la detección del estado emocional, analizando las correspondencias y los posibles patrones existentes entre los cambios fisiológicos y la emoción percibida por el mismo.

Existen estudios que afirman que la actividad fisiológica del sujeto representa un componente importante de su estado afectivo [159]. Diversos trabajos [106, 148, 123, 125, 186, 95, 94] han demostrado que existen correlaciones entre el estado emocional y las mediciones y características extraídas de señales como la variabilidad del pulso cardíaco (HRV o *Heart Rate Variability*), la respuesta galvánica de la piel (GSR o *Galvanic Skin Response*), la temperatura corporal, los patrones de respiración o las bandas de las señales procedentes de electroencefalografía (EEG).

Desde el punto de vista de la predicción del estado afectivo del usuario, una de las principales ventajas que aporta el estudio de señales fisiológicas frente al análisis de expresiones faciales es que pueden proporcionar información directa del sujeto difícilmente falsificable, en contraposición a las expresiones fingidas que pueden mostrar un estado afectivo diferente del percibido por el sujeto o incluso inhibirlo. Medidas del sistema nervioso central como las ondas cerebrales procedentes de un EEG o del sistema nervioso periférico como el pulso cardíaco, GSR o la temperatura de la piel, son difícilmente controlables por una persona deliberadamente y, sin embargo, todas pueden aportar información valiosa sobre su estado afectivo [164].

Por el contrario, el principal inconveniente que presenta el estudio de este tipo de señales es que su obtención resulta más invasiva que la procedente de otras fuentes de información, aspecto que dificulta su aplicación en entornos de interacción reales, aunque la aparición de dispositivos inalámbricos con micro y nano sensores integrados en *wereables* [132], en *e-textiles* [69], en dispositivos de entrada como el ratón o incluso en la silla, pueden minimizar el impacto. Otros inconvenientes que suelen presentar estas señales son: 1) su tratamiento suele ser computacionalmente más complejo y costoso, especialmente en el caso de señales procedentes de EEG por su alta resolución; 2) son sistemas completamente dependientes de un único usuario, lo que impide el tratamiento de señales procedentes de varios sujetos simultáneamente; 3) son susceptibles de verse afectadas por ruidos generados por los elementos electrónicos que capturan las señales, por movimientos involuntarios del usuario como consecuencia de la reacción a una emoción o simplemente por una mala colocación de los sensores; 4) el análisis de la información contenida en este tipo de señales suele requerir una cantidad de información relativamente larga en el tiempo para poder tomar decisiones sobre el estado afectivo del usuario.

En lo referente a las señales fisiológicas adquiridas a través de la actividad electrodermal, la respuesta galvánica de la piel (GSR) se define como los cambios en las propiedades eléctricas de la piel como consecuencia de la sudoración. Suele representar un indicador del nivel de excitación sufrido como consecuencia de un estímulo externo [125, 133]. Por otro lado, las variaciones en la temperatura corporal se encuentran correlacionadas con cambios en el flujo sanguíneo como consecuencia de la resistencia arterial a la presión sanguínea modulada por la tensión muscular. Estos cambios de temperatura pueden también reflejar variaciones en el estado emocional del sujeto, especialmente los cambios de temperatura producidos en las manos, de modo que una emoción positiva suele estar asociada a aumentos de temperatura, mientras que una negativa a descensos de la misma [151].

La electromiografía (EMG) mide la actividad muscular del sujeto (contracciones) y se encuentra relacionada con estados afectivos negativos como Miedo o Estrés [129, 133], mientras que los electrooculogramas (EOG) consisten en la medición del movimiento de los ojos y de sus efectos, como por ejemplo los producidos por el parpadeo, mediante la colocación de pequeños electrodos en los alrededores de los músculos oculares. No obstante, aunque el análisis de patrones EOG puede resultar de interés en diversas áreas, en el campo de la computación afectiva no ha sido extensivamente utilizado, aunque suele resultar de utilidad para la eliminación de posibles interferencias producidas por los movimientos oculares durante la adquisición de señales procedentes de EEG [25].

En [85] se reporta cómo afectan los cambios emocionales en las variaciones de la presión sanguínea, independientemente de la postura o ubicación del sujeto durante su medición, afirmando que estados afectivos asociados a la excitación (Ira, Ansiedad y Felicidad) producen incrementos significativos en la presión sistólica y diastólica. En dicho estudio se concluye que la presión sanguínea es mayor en estados los afectivos asociados a la Ira o la Ansiedad en comparación con los estados asociados a la Felicidad.

El análisis de los patrones respiratorios del sujeto puede también aportar

información sobre su estado afectivo [148]. Una respiración lenta se encuentra relacionada con estados de relajación, mientras que una respiración irregular con variaciones rápidas. Las arritmias respiratorias pueden estar relacionadas con emociones como la Ira o el Miedo [97].

Las señales procedentes de electrocardiogramas (ECG) han sido extensivamente utilizadas en el estudio del estado afectivo del usuario [95, 94, 97]. Con ellas se obtienen diferentes parámetros de la actividad cardíaca como el ritmo cardíaco (HR), para diferenciar entre emociones positivas y negativas [106] o la variabilidad del ritmo cardíaco (HRV), como indicación del esfuerzo mental y el estrés [94]. Además, las señales cardíacas presentan una fuerte correlación con los patrones de respiración en la manifestación de un estado afectivo [148].

El estudio de señales del sistema nervioso central procedentes de EEG y su asociación con estados afectivos es una tarea compleja. Zheng et al. [194] realizan una interesante y completa revisión de los trabajos más recientes relacionados con el análisis de señales EEG y sus emociones. La complejidad de la correlación entre este tipo de señales y las emociones subyacentes se debe principalmente a las variaciones en los patrones EEG inter-sujetos como respuesta a una misma emoción, aspecto que dificulta la definición de patrones emocionales estables en el tiempo, necesarios para su aplicación en entornos reales, aunque en este sentido se están realizado progresos [103, 193, 194]. Adicionalmente, este tipo de señales presenta el problema de su alta resolución y, por consiguiente, la cantidad de información disponible por unidad de tiempo para su posterior procesamiento (considérese, por ejemplo, un sistema de 32 electrodos que capturan información a razón de 256 Hz durante unos minutos). Por este motivo, en el análisis de señales EEG es habitual extraer y analizar las diferentes bandas que las componen en función de la frecuencia de la onda: δ para frecuencias inferiores a 4 Hz; θ en el rango 4-7 Hz; α entre 8-13 Hz; β entre 14-30 Hz; y γ para frecuencias superiores a 31 Hz. Por otra parte, debido a que existen evidencias de que la lateralización entre hemisferios puede estar asociado con las emociones [47], también es frecuente obtener características de las señales EEG entre pares de electrodos simétricos a ambos lados del cerebro, con el propósito de estudiar las posibles asimetrías existentes entre ambos hemisferios [169, 97].

Para la implementación y entrenamiento de sistemas afectivos basados en señales de carácter fisiológico es necesario contar con bases de datos anotadas y de acceso público. A este respecto, entre los diferentes corpus que contienen información fisiológica, específicamente señales procedentes de EEG, destacan MAHNOB-HCI [169], DEAP (*Database for Emotion Analysis using Physiological Signals*) [97], EMDB (*Emotional Movie Database*) [34] y DECAF (*Multimodal Dataset for Decoding Affective Physiological Responses*) [1]. Por otro lado, en lo referente al estudio exclusivo de señales EEG obtenidas de sujetos en diferentes sesiones se encuentra SEED (*SJTU Emotion EEG Dataset*) [193]. En esta base de datos cada participante fue grabado tres veces en intervalos de una semana y puede resultar de interés para analizar la correspondencia entre patrones EEG estables en el tiempo y sus correspondientes emociones.

Como se expuso anteriormente, emociones y estados afectivos de relajación o excitación producen cambios significativos en las respuestas fisiológicas de las

personas. Estos cambios se manifiestan con variaciones en la presión sanguínea, en el ritmo cardíaco, en los patrones respiratorios o en la temperatura o sudoración de la piel. En este sentido, cuando una emoción emerge en una persona, por ejemplo Ira, ésta suele venir acompañada de una expresión facial determinada junto con unos cambios fisiológicos como el incremento del pulso cardíaco. Con estas evidencias es natural pensar que es posible determinar con mayor precisión el estado afectivo de un sujeto combinando diferentes observaciones [113], es decir empleando un enfoque multimodal en contraposición a los enfoques unimodales basados en una única fuente de información. En lo referente a la combinación o fusión de características fisiológicas, Calvo y D'Mello [32] resumen algunos de los principales estudios donde han sido combinadas diferentes señales para la determinación del estado emocional del usuario, tales como ECG, EMG, GSR, temperatura de la piel, patrones de respiración o EEG.

Por último, es destacable la extensa y actualizada revisión que realizan D'Mello y Kory [55] sobre sistemas afectivos multimodales no circunscrita exclusivamente al ámbito de señales fisiológicas, evaluando, adicionalmente, las mejoras de los sistemas multimodales frente a los unimodales.

2.2.3. Patrones de comportamiento

Es evidente que existe una relación entre el comportamiento de un usuario cuando interactúa con un sistema y su estado afectivo y mental subyacente. Por este motivo, la identificación de patrones de comportamiento asociados a emociones concretas puede ayudar en el diseño de sistemas inteligentes y adaptativos que ayuden al usuario a maximizar su experiencia de usuario anticipándose o adaptándose a sus necesidades, de modo que pueda interactuar con el sistema de un modo más satisfactorio y eficaz en función de su estado afectivo.

Las técnicas basadas en patrones de comportamiento son ampliamente utilizadas en sistemas basados en entornos inteligentes (*ambient intelligent*) [10]. En estos contextos la información es recabada mediante diferentes sensores con el objetivo de aprender los hábitos y las preferencias de los usuarios, con la intención de prestarle proactivamente servicios personalizados.

Entre los métodos para la recolección de la información sobre cómo interactúa un usuario con el sistema para poder determinar, *a posteriori*, si existen patrones de comportamiento asociados a un estado afectivo concreto, se encuentran los basados en técnicas de visión, análisis de la voz, señales fisiológicas, registros de interacción usuario-software basados en los movimientos, velocidad y pulsaciones del ratón, del teclado o de ambos [195, 88, 185, 65, 99, 23, 157], registros de tiempos de respuesta, tiempo consumido en la resolución de una tarea, tasas de errores y aciertos cometidos, número y tipo de ayudas solicitadas, así como cuestionarios y auto-evaluaciones para reportar sus percepciones y sentimientos durante su interacción con el sistema como, por ejemplo, mediante formularios SAM [27] que recojan su percepción en términos de Valencia, Activación y Dominancia [7, 162].

Los test SAM son formularios exclusivamente pictóricos para que el usuario puntúe en una escala de cinco valores, nueve considerando las posibles puntuaciones entre valores intermedios, su estado emocional como respuesta a la reacción

de un estímulo. La auto-evaluación se realiza sobre las dimensiones afectivas correspondientes a Valencia, Activación y Dominancia. Debido a su diseño no verbal, este tipo de test es idóneo para su utilización en cualquier ámbito, edad del usuario o entorno cultural.

Independientemente del método de recolección de la información, ésta tendrá que ser analizada para poder crear sistemas que sean capaces de tomar decisiones en función de los patrones de comportamiento descubiertos sobre la información recogida. Si dicha información proviene de varias fuentes independientes, éstas necesitarán ser previamente preprocesadas, sincronizadas y, en su caso, combinadas convenientemente. Las técnicas de aprendizaje automático (*machine learning*) son las más utilizadas para el análisis de patrones, aunque el uso de estereotipos se ha sugerido en algunos estudios como una herramienta para acelerar el proceso de aprendizaje inicial [98].

En lo referente a la disponibilidad de bases de datos públicas para el el diseño y entrenamiento de sistemas afectivos basados en patrones de comportamiento, no se han encontrado referencias a corpus de este tipo debido, principalmente, a que estos sistemas son altamente dependientes del dominio y del contexto en el que se ejecutan y, por tanto, difícilmente extrapolables a otras aplicaciones diferentes para las que fueron diseñadas.

2.3. Conclusiones

La detección automática del estado emocional de un usuario es un área de interés creciente en diversidad de entornos como el publicitario, el de la salud, los videojuegos o la educación, entre otros. El objetivo principal de los sistemas que implementan métodos de reconocimiento afectivo es adaptar la respuesta del sistema para que la percepción del usuario en su interacción sea más amigable, efectiva, precisa y productiva.

No obstante, como se ha podido extraer de lo expuesto en este capítulo, su detección, análisis y respuesta es una tarea compleja, no exenta de imprecisiones y limitada cuando se compara con la percepción realizada por un humano.

En la primera parte de este capítulo se ha analizado el impacto que las emociones pueden suponer en el aprendizaje desde el punto de vista de los sistemas adaptativos basados en la tutorización inteligente.

Un STI que incorpore soporte afectivo podría inferir en cada momento de la resolución de un problema el estado afectivo del estudiante y adaptar su funcionamiento al mismo. Esta información puede ser combinada con el modelo cognitivo del usuario que el STI haya ido conformando durante su interacción, junto con información histórica proveniente de otros alumnos (ficheros de *log*), la dificultad media del problema que se está realizando, autoevaluaciones del usuario sobre su estado afectivo tras la resolución de un problema así como con cualquier otra variable que pudiera resultar de utilidad, con el objetivo final de aplicar las estrategias instruccionales que permitan regular las dimensiones afectivas del estudiante, minimizando situaciones de aburrimiento, frustración o abandono y, por tanto, mejorar su aprendizaje.

En la segunda parte se ha realizado una revisión de las técnicas y métodos habitualmente utilizados en la detección automática del estado emocional del usuario desde tres perspectivas diferenciadas y no excluyentes: mediante visión por computador; a través del análisis de las señales fisiológicas del sujeto; y por último, la menos extendida pero no así menos interesante, mediante el análisis y la relación entre posibles patrones de comportamiento asociados a estados afectivos.

Las técnicas basadas en visión suelen centrar su atención en el análisis del rostro mediante modelos de representación de la cara en 2D o 3D. Han sido ampliamente implementadas en sistemas de reconocimiento afectivo, principalmente por dos motivos: por imitación humana y por su baja intrusividad. No obstante, presentan varios problemas que dificultan el reconocimiento de la emoción mediante estas técnicas, como son los posibles cambios en la iluminación, rotaciones de cabeza, oclusiones o diferencias entre sujetos (cambios en la textura de la piel, rasgos étnicos o de la edad).

Los métodos basados en el análisis de señales fisiológicas se encargan de analizar variables biométricas como el ritmo cardíaco, la temperatura corporal, la conductividad de la piel, EEG, etc. Presentan el problema de que son más intrusivos que los basados en visión, requieren un instrumental más complejo y preciso, así como un tratamiento computacionalmente más costoso. Como ventaja aportan que pueden proporcionar información directa del sujeto difícilmente falsificable por el mismo.

Por último, los patrones de comportamiento intentan dar respuesta al modo de interacción de un usuario con el sistema, aprendiendo sus hábitos y preferencias para poder prestarle proactivamente ayudas o servicios personalizados. Su principal problema es que sus aplicaciones son altamente dependientes del dominio y, por tanto, difícilmente extrapolables a otros contextos.

No obstante, todas estas técnicas y métodos son susceptibles de ser integrados mediante diversas estrategias de fusión dando lugar a sistemas de análisis multimodales. Estos sistemas presentan, en general, un rendimiento mejor que sus equivalentes unimodales.

Capítulo 3
Aprendizaje automático

Resumen
En este capítulo se describen las principales técnicas y métodos de aprendizaje automático empleados en los trabajos y aportaciones desarrolladas a lo largo de esta tesis para la detección del estado emocional de un usuario, así como los procedimientos para la evaluación de su rendimiento.

Contenidos

3.1. Clasificación y métodos 26
3.2. Regresión 32
3.3. Reducción de la dimensionalidad sobre imágenes 34
3.4. Análisis de series temporales: DTW 36
3.5. Evaluación del rendimiento de los clasificadores 37

El aprendizaje automático (*machine learning*) es una disciplina perteneciente a las ciencias de la computación y a la inteligencia artificial, cuyo objetivo es la creación de sistemas con capacidad para aprender y actuar en una tarea o conjunto de tareas determinadas sin la necesidad de una programación explícita sobre el conocimiento con el que operan (basado éste exclusivamente en su experiencia), de modo similar a como lo haría un humano.

El proceso de aprendizaje puede ser supervisado o no supervisado. Se considera supervisado cuando las muestras utilizadas tanto en la fase de entrenamiento como en la fase de test o evaluación se encuentran previamente etiquetadas en clases. El aprendizaje no supervisado, por el contrario, no dispone de un etiquetado de las muestras. En este caso, se cuenta con un conjunto de muestras con unas características determinadas que pueden ser agrupadas en clases atendiendo a ciertos criterios de similitud de características (técnicas de *clustering*).

En este contexto, cuando se plantea la necesidad de desarrollar un sistema capaz de detectar, predecir o determinar el estado emocional de sus usuarios,

independientemente del número de emociones, estados afectivos o de expresiones faciales que se pretendan reconocer, es necesario implementar un procedimiento que asigne cada una de las muestras de entrada a una clase, emoción o estado afectivo determinado de salida. Este proceso de categorización en clases es lo que se conoce habitualmente como clasificación. En la figura 3.1 se muestra un diagrama del proceso de clasificación para N posibles clases.

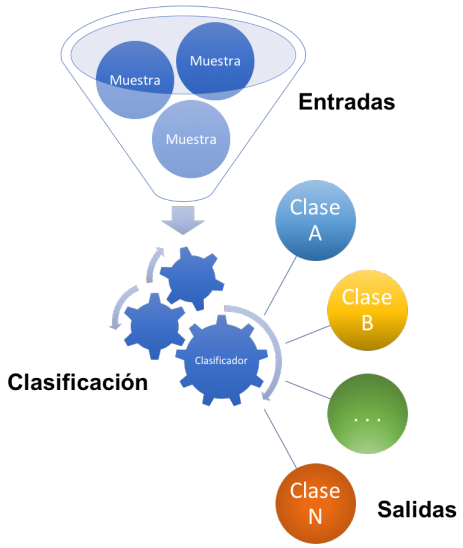


Figura 3.1: Diagrama genérico del proceso de clasificación

En los siguientes apartados se introducirán los fundamentos teóricos de los métodos y algoritmos de aprendizaje automático empleados en los trabajos que se desarrollarán en la Parte II (Aportaciones en la detección emocional) y Parte III (Aportaciones en la mejora instruccional) de este documento. En concreto, se describirán algunas de las técnicas habituales de reducción de la dimensionalidad de una imagen como proceso previo a la clasificación, como lo es PCA (Análisis de componentes principales) y LDA (Análisis discriminante lineal), algunos de los algoritmos de clasificación clásicos como el vecino más próximo (K-NN), árboles de decisión C4.5 o LMT (Árboles de regresión logística), así como algunos de los métodos tradicionales para la identificación de patrones complejos como son las máquinas de soporte vectorial (SVM).

3.1. Clasificación y métodos

3.1.1. Vecino más próximo

El algoritmo del vecino más próximo o K-NN (*K Nearest Neighbor* en inglés) [44] es un método de clasificación local no paramétrico (no tiene en cuenta la distribución de las muestras) basado en el cálculo de la distancia mínima.

Este método permite inferir la categoría de una nueva muestra atendiendo al grado de similitud o proximidad mediante la comparación con las K muestras más próximas (vecinos), seleccionando, de entre ellas, la clase con mayor número de muestras. La medida comúnmente empleada en el cálculo de la proximidad a los vecinos es la distancia euclídea, aunque cualquier otra medida puede ser aplicada.

Es decir, dada una muestra x a clasificar entre sus k vecinos más próximos $y_1, y_2, \ldots, y_k$, la clase $c(x)$ escogida con mayor número de muestras será determinada mediante la ecuación

$$c(x) = \arg\max_{c \in C} \sum_{i=1}^{k} \delta(c, c(y_i)) \tag{3.1}$$

donde $c(y_i)$ es la clase de pertenencia del vecino y_i y δ la función de distancia empleada para el cálculo de la proximidad.

En la figura 3.2 se muestra un ejemplo para la clasificación de una nueva muestra x mediante el algoritmo K-NN para un valor $K = 3$. En este caso la nueva muestra x será clasificada como un ejemplo de clase A.

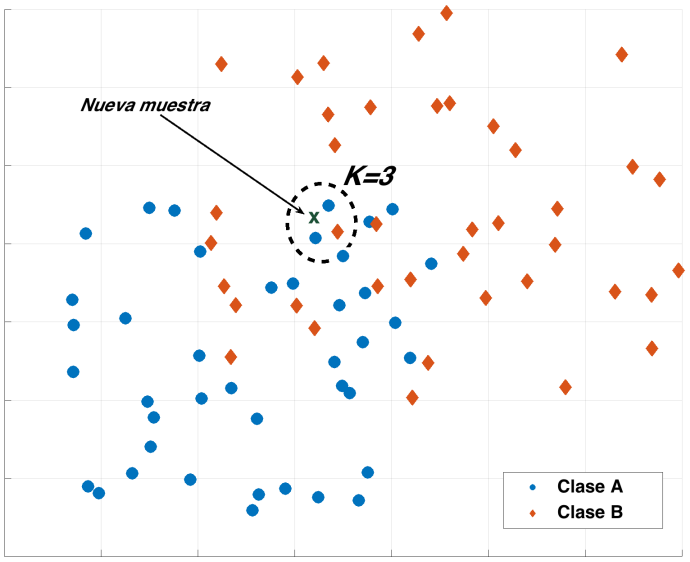


Figura 3.2: Ejemplo de la clasificación de una muestra mediante el algoritmo del vecino más próximo con $K = 3$

3.1.2. Árboles de decisión

En aprendizaje automático supervisado, los árboles de decisión son utilizados como modelos predictivos en los que a partir de un conjunto de muestras u observaciones (conjunto de entrenamiento) se construye, mediante una serie de

grafos dirigidos, un conjunto de nodos que inducen las categorías a las que pertenecen las muestras en función de sus características.

Los árboles de decisión se componen de nodos y aristas. Los nodos intermedios o de decisión representan los diferentes atributos de la muestra, mientras que los nodos finales u hojas las posibles categorías. Las aristas relacionan los nodos con los posibles valores que éstos pueden tomar. Un ejemplo de un árbol de decisión para la clasificación de un vehículo entre sus posibles categorías se muestra en la figura 3.3.

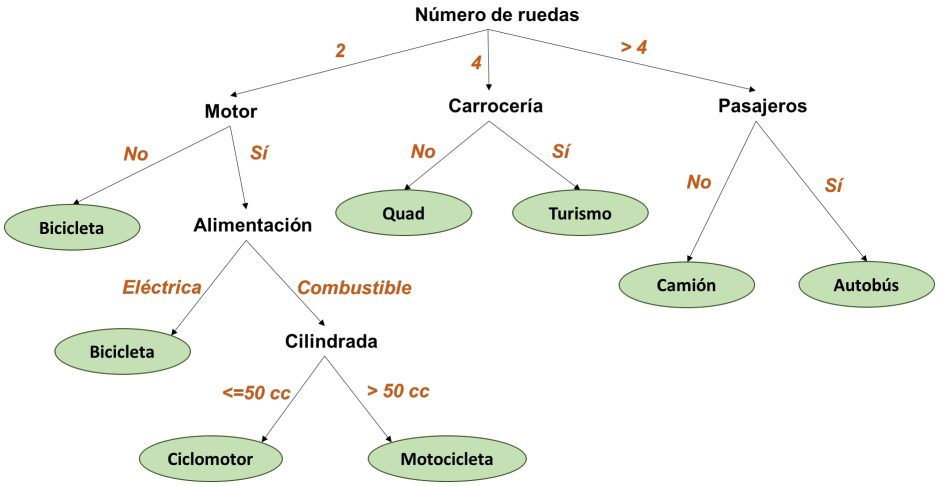


Figura 3.3: Ejemplo de clasificación de un vehículo entre sus posibles categorías mediante un árbol de decisión

Uno de los algoritmos más populares en la generación de árboles de decisión es ID3 (*Iterative Dichotomiser 3* en inglés), definido por Quinlan [146]. Este algoritmo construye el árbol seleccionando iterativamente cada atributo en función de la entropía de los datos tras la selección, escogiendo aquel cuya entropía es la menor y reduciendo el problema a subárboles de menor complejidad mediante una estrategia descendente de tipo "divide y vencerás".

Entre las ventajas que aporta el algoritmo ID3 se encuentran: 1) la construcción y evaluación del árbol es simple y rápida; 2) los árboles construidos suelen ser poco profundos; 3) el proceso de clasificación se detiene al llegar a un nodo final (categoría), aspecto que limita el número de iteraciones y, por tanto, el tiempo de cálculo. Por otro lado, el algoritmo presenta una serie de limitaciones: 1) durante el proceso de clasificación y recorrido del árbol únicamente se comprueba un atributo en cada uno de los pasos. Esto puede provocar que se alcance un mínimo local al no disponer de mecanismos de retroceso; 2) incapacidad para trabajar con datos incompletos; 3) la clasificación de datos continuos puede resultar costosa por la cantidad de subárboles a crear: 4) puede presentar problemas de sobreajuste (*overfitting* en inglés).

Para salvar las limitaciones de ID3 Quinlan creó una variante a su algoritmo

original denominada C4.5 (o J48 en algunos paquetes de minería de datos) [147]. Este algoritmo incorpora: 1) capacidad para tratar con atributos con valores continuos mediante la definición de un umbral y la división de las muestras en función del umbral; 2) un mecanismo de poda mediante el cual se permite la sustitución de una parte del árbol construido por un nodo terminal que reduzca el error del subárbol sobre el conjunto de prueba, resolviendo de esta manera posibles problemas de sobreajuste; 3) posibilidad de tratar con valores incompletos; 4) asignación de pesos a los atributos.

Por otro lado, los árboles LMT o de regresión logística (*Logistic Model Trees* en inglés) son árboles de inducción cuyas nodos terminales contienen funciones de regresión logística. Es decir, combinan dos métodos de aprendizaje complementarios: los árboles de inducción y la regresión logística. Estos árboles han demostrado, en general, una precisión mayor que los implementados mediante el algoritmo C4.5 [105].

3.1.3. Perceptrón multicapa

El perceptrón multicapa [152] es un tipo de red de neuronas artificiales usado con frecuencia en problemas de aprendizaje supervisado por su simplicidad y por su habilidad como aproximador universal. Este modelo se encuentra basado en otro menos complejo denominado, simplemente, perceptrón.

La arquitectura del perceptrón se fundamenta en una capa de neuronas artificiales con pesos sinápticos w_i que ponderan las entradas x_i y un umbral u, ambos parámetros ajustables. En la figura 3.4 se muestra una representación de la estructura típica de un perceptrón.

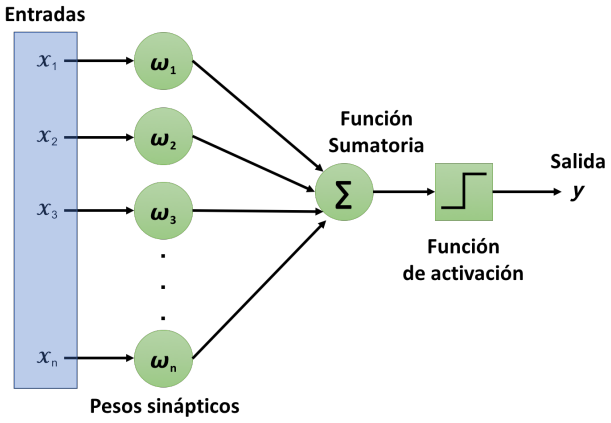


Figura 3.4: Estructura típica de un perceptrón

La salida y o función de transferencia del perceptrón se puede formular como

$$y = f(x_1, x_2, x_3, \ldots, x_n) = \begin{cases} 1 & \text{si } wx_1 + wx_2 + wx_3 + \ldots + wx_m \geq u \\ 0 & \text{si } wx_1 + wx_2 + wx_3 + \ldots + wx_m < u \end{cases} \quad (3.2)$$

donde la función de activación se ha definido como la función de *Heaviside*, una de las más comúnmente utilizadas. Otras funciones de activación habituales son la sigmoidal y la tangente hiperbólica.

De modo general, el proceso de aprendizaje del perceptrón consiste en un método de error y corrección mediante el cual se realiza un ajuste en los pesos w_i y en el umbral u mediante un proceso adaptativo que compara, iterativamente, la salida del perceptrón con el resultado esperado, ajustando sus parámetros en función del error obtenido.

El perceptrón multicapa es una extensión del perceptrón simple, desarrollado por las limitaciones que presentaba este último para resolución de problemas no lineales [130]. A diferencia del perceptrón simple, el perceptrón multicapa se compone de un número arbitrario l de capas de neuronas artificiales ocultas. La figura 3.5 muestra una representación genérica de la arquitectura típica de un perceptrón multicapa.

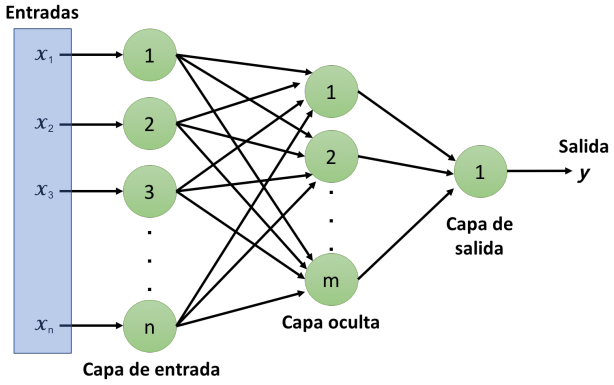


Figura 3.5: Estructura genérica de un perceptrón multicapa

El flujo de información de entrada al perceptrón multicapa se propaga hacia delante, desde su capa de entrada, pasando por las capas ocultas, hasta la salida. El mecanismo de aprendizaje usado en el perceptrón es de retropropagación del error (o regla *delta*). En esta fase de aprendizaje, de modo similar a como lo hace el perceptrón, la red ajusta los pesos y umbrales de las neuronas para minimizar el error en la predicción.

3.1.4. Máquinas de soporte vectorial

Las máquinas de soporte vectorial [43] (SVM o *Support Vector Machines* en inglés) constituyen un método de aprendizaje de clasificación lineal, cuyo propósito es la categorización de las muestras realizando una separación del espacio muestral mediante un hiperplano que maximice la distancia mínima entre la proyección ortogonal de las muestras y el hiperplano. Esta región de distancias mínimas, comúnmente denominada margen geométrico o margen máximo, es la que define el hiperplano óptimo entre los infinitos hiperplanos de separación que pueden existir.

El conjunto de muestras de entrenamiento de cada clase que se encuentran en la frontera de los márgenes geométricos se denominan vectores soporte.

En la figura 3.6 se muestra un ejemplo del hiperplano óptimo obtenido para la separación lineal de unas muestras pertenecientes a dos posibles clases distribuidas en un espacio bidimensional, el margen geométrico y los vectores soporte situados en las fronteras del margen.

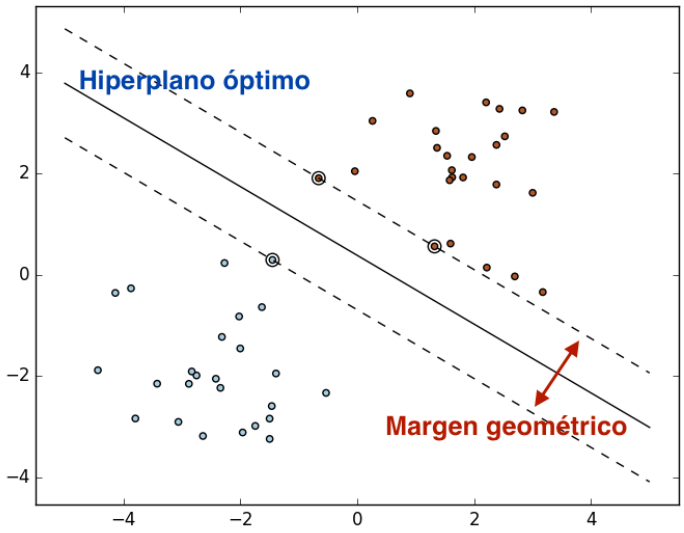


Figura 3.6: Ejemplo del hiperplano óptimo definido para la separación de dos clases distribuidas en un espacio bidimensional

Cuando las muestras se distribuyen de forma que no es posible realizar una separación lineal, la solución consiste en la aplicación de una transformación no lineal a un espacio de características de mayor dimensionalidad linealmente separable. Estas transformaciones son realizadas a través de funciones denominadas *kernel*, así como del ajuste de los parámetros que las definen. Algunas funciones *kernel* comúnmente empleadas son: lineal, polinómica, gaussiana (RBF) o sigmoide.

No obstante, es frecuente encontrar situaciones en las que no es posible aplicar sobre los datos una transformación no lineal que permita realizar una separación lineal. En estos casos la estrategia adoptada consiste en la asunción de que no será posible separar correctamente todas las muestras de entrenamiento y que, por tanto, existirán algunos errores en la clasificación. Estos errores pueden ser controlados mediante un parámetro de coste C que permitirá determinar en qué medida influirá el impacto de los ejemplos no separables en la definición del hiperplano. Valores elevados para el parámetro C implica muestras ampliamente separables con riesgo a realizar un sobreajuste (o memorización del problema) en el clasificador sobre las muestras de entrenamiento, mientras que valores pequeños supone admitir la existencia de un número elevado de muestras de entrenamiento no separables. En este sentido, la elección de un valor adecuado de C supone un

compromiso entre el riesgo de sobreajuste y el número de muestras no separables. Su estimación suele realizarse heurísticamente durante el entrenamiento de la SVM o mediante técnicas de búsqueda exhaustiva de tipo *Grid Search*.

Las SVM se pueden generalizar para realizar la clasificación de las muestras mediante regresión (lineal o no lineal), con fundamentos similares a los descritos para las SVM. Estos sistemas son conocidos como SVR (*Support Vector Regression machines* en inglés).

3.2. Regresión

La regresión es una generalización del método de clasificación en el que la salida del sistema de aprendizaje consiste en un valor o conjunto de valores continuos, en contraposición a una salida categórica y discreta.

El método de regresión consiste en la aproximación de una función $f(X) = \hat{Y}$ que mejor describa el comportamiento de una variable Y en función de sus entradas X. La regresión será lineal si la función de ajuste empleada sobre los datos es lineal. Por el contrario, la regresión será no lineal si la función de ajuste que la aproxima tampoco lo es.

De forma genérica, un modelo de regresión lineal $\hat{Y}$ que aproxima la relación entre la variable dependiente Y con las k variables independientes X se define como

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i X_i + \epsilon \tag{3.3}$$

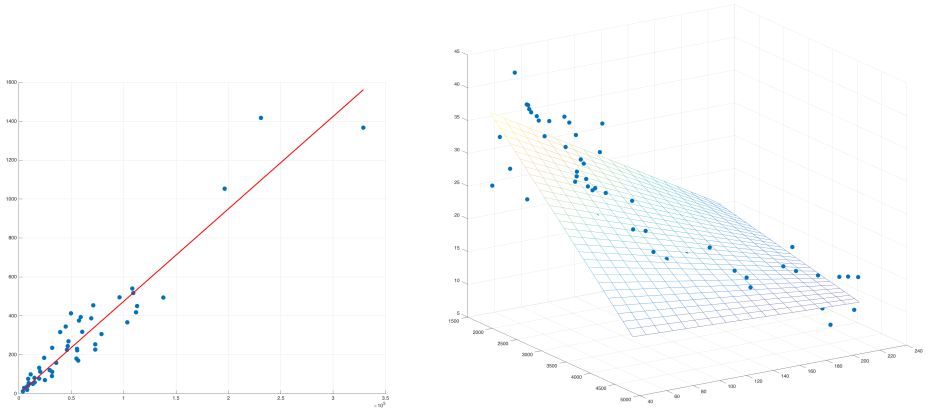
donde $\hat{\beta}_i$ representa los coeficientes de regresión estimados que definen la influencia de cada variable independiente y ϵ el error de la aproximación.

La regresión lineal definirá la función de regresión mediante una recta en el caso de una regresión lineal simple (interviene únicamente una sola variable independiente) o un plano o un hiperplano si la regresión lineal es múltiple (dos o más variables independientes, respectivamente).

En la figura 3.7(a) se muestra un ejemplo en el que se aproxima el comportamiento de una variable dependiente con respecto otra independiente a través de una recta obtenida mediante regresión lineal simple, mientras que en la figura 3.7(b) se muestra otro ejemplo con dos variables independientes y el plano obtenido mediante regresión lineal múltiple.

Es posible que la regresión lineal no sea capaz de modelar adecuadamente la relación entre la variable dependiente y las independientes. En estos casos, la regresión no lineal puede proporcionar una solución mediante la aproximación de una función no lineal (polinómica de grado n, logarítmica, exponencial, etc.). En la figura 3.8 se muestra un ejemplo de aproximación mediante regresión no lineal.

Dado que la regresión representa una aproximación al comportamiento de un sistema, es posible medir la bondad del ajuste del modelo regresivo con respecto al sistema real modelado mediante medidas de dispersión del error entre el modelo y los datos. En este sentido, las mediciones comúnmente empleadas son el coeficiente



(a) *Regresión lineal simple* **(b)** *Regresión lineal múltiple*

Figura 3.7: Ejemplos de regresión lineal

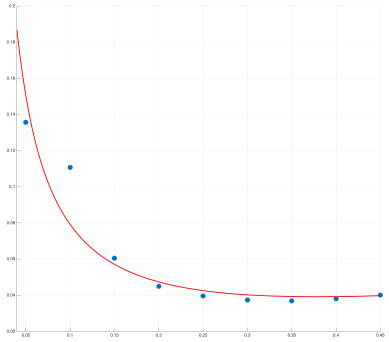


Figura 3.8: Ejemplo de regresion no lineal

de determinación R^2 y la raíz del error cuadrático medio $RMSE$ (*Root Mean Square Error* en inglés).

El coeficiente de determinación R^2 mide la capacidad explicativa de un modelo de regresión lineal como la proporción de variación total de la variable dependiente Y respecto a su media, con valores acotados por $0 \leq R^2 \leq 1$. Un valor de R^2 próximo a la unidad indica que el ajuste del modelo es ideal, es decir que la variación total del sistema es explicada por el modelo regresivo, mientras que un valor próximo a cero indica lo contrario, que el modelo regresivo no representa al sistema real que intenta modelar.

El coeficiente de determinación R^2 se define como

$$R^2 = 1 - \frac{SSE}{TSS} \tag{3.4}$$

siendo SSE la suma de los cuadrados del error residual (*Sum of Squares Error* en

inglés) definida como

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.5}$$

y TSS la suma de los cuadrados totales (*Total Sum of Squares* en inglés), definida como

$$TSS = \sum_{i=1}^{n}(y_i - \overline{y})^2 \tag{3.6}$$

donde $\overline{y}$ representa la media de las observaciones y_i.

No obstante, R^2 puede proporcionar una visión poco precisa del ajuste del modelo regresivo al no tomar en cuenta el tamaño de la muestra. En este sentido, el coeficiente de determinación ajustado $\bar{R}^2$, o simplemente R^2 ajustado, toma en consideración el tamaño muestral debido, principalmente, a que R^2 y el número de datos suelen ser inversamente proporcionales, de modo que un número reducido de observaciones podría proporcionar un valor próximo a la unidad para R^2 sin que necesariamente exista una alta relación lineal entre las variables.

El coeficiente de determinación ajustado $\bar{R}^2$ se define como

$$\bar{R}^2 = 1 - \frac{N-1}{N-k-1}(1-R^2) \tag{3.7}$$

siendo N el tamaño de la muestra y k el número de variables.

Por otro lado, el $RMSE$ se define como la desviación estándar del error en las predicciones realizadas por el modelo regresivo $\hat{Y}$ con respecto al modelo original Y.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.8}$$

A diferencia del coeficiente de determinación R^2 cuyos valores se encuentran restringidos al rango [0–1], los valores de $RMSE$ no se encuentran limitados a ningún rango determinado. Por esto, aunque pueda parecer que R^2 es más fácilmente interpretable, $RMSE$, sin embargo, es una medida capaz de explicitar mejor la desviación de las predicciones.

Por último, la regresión logística es un tipo particular de regresión utilizado en sistemas de aprendizaje cuyo propósito es predecir un conjunto de variables categóricas, en contraposición a la regresión clásica cuya salida consiste en un valor o conjunto de valores continuos.

3.3. Reducción de la dimensionalidad sobre imágenes

Cualquier imagen bidimensional de $m \times p = n$ píxeles puede ser representada como un único punto en un espacio de características n-dimensional, simplemente

mediante la concatenación de sus m filas o p columnas.

Consideremos un conjunto de s clases $C = \{c_1, c_2, \ldots, c_s\}$, donde cada c_i representa una expresión facial; y por otro lado un conjunto compuesto de las muestras etiquetadas $X = \{(x_1, l_1), (x_2, l_2), \ldots, (x_p, l_p)\}$, siendo x_i una imagen en el espacio de alta dimensionalidad (el conjunto de entrenamiento), y $l_i \in C$ una etiqueta que representa su expresión facial. Asumamos otro conjunto de q imágenes sin etiquetar $E = \{e_1, e_2, \ldots, e_q\}$, siendo e_i una imagen en el espacio de alta dimensionalidad (conjunto de test).

Con este planteamiento, la clasificación puede ser concebida como un método supervisado para la identificación de una expresión facial en el espacio n-dimensional para cada muestra de E.

No obstante, es previsible que tras proyección de las imágenes éstas no se encuentren distribuidas de forma aleatoria en este espacio de alta dimensionalidad, sino que, por el contrario, existan regularidades estadísticas que hagan posible construir un subespacio de imágenes de caras con una dimensionalidad significativamente menor.

En esta línea, un método no supervisado habitualmente empleado por su simplicidad para reducir la alta dimensionalidad de este espacio es PCA [73, 81]. El método se usa con el objetivo de encontrar fuertes correlaciones entre los datos que permitan reducir su dimensionalidad y establecer los vectores que mejor se ajustan a la distribución de los elementos X en el espacio original de la imagen (componentes principales). Estos vectores corresponden a las direcciones de varianza máxima en el espacio original de la imagen, definiendo, mediante un conjunto de vectores propios o *eigenvectores*, un subespacio con una dimensionalidad inferior al espacio original, pero reteniendo gran parte de la información.

Formalmente, la distribución de la imagen en su espacio original se puede definir como

$$S = \sum_{k=1}^{p} (x_k - \bar{x})(x_k - \bar{x})^T \tag{3.9}$$

donde $\bar{x}$ representa la media de todas las muestras. A partir de ella, PCA calcula la transformación lineal W^T que maximiza la distribución de la proyección de las muestras $W^T SW$. Esta matriz de transformación permite la proyección de cualquier muestra en el espacio original de características n-dimensional en un vector en el nuevo espacio m-dimensional. Con este nuevo espacio, la clasificación de una nueva muestra se llevará a cabo en el espacio proyectado utilizando, por ejemplo, un clasificador basado en una estrategia de tipo "vecino más próximo" usando para ello las etiquetas del conjunto de entrenamiento en X.

Este enfoque de reducción de la dimensionalidad mediante PCA fue aplicado con éxito en un método para el reconocimiento de rostros denominado *Eigenfaces* [182]. Para reducir este espacio, el método *Eigenfaces* obtiene mediante PCA una transformación lineal desde el espacio de características n-dimensional de la imagen de la cara a un espacio de características m-dimensional, con $m \ll n$, denominado como "face space". Este método es susceptible de ser extendido mediante múltiples subespacios para el reconocimiento de expresiones faciales.

Otro método comúnmente utilizado para problemas en los que se desea reducir la dimensionalidad de un conjunto de datos es el Análisis Discriminante Lineal (LDA) [68]. Este método, de modo equivalente a PCA, se fundamenta en una transformación lineal de los datos para la reducción de la dimensionalidad, con la particularidad de que mientras PCA obtiene los vectores correspondientes a las direcciones de los vectores de varianza máxima en el espacio original de los datos sin tener en cuenta las etiquetas de las muestras, LDA calcula las direcciones que maximizan (discriminan) la separación entre clases. Es decir, mientras que PCA es un método de reducción no supervisado, LDA es considerado como supervisado.

3.4. Análisis de series temporales: DTW

Un método comúnmente empleado en el análisis de secuencias de valores a lo largo de un eje temporal es DTW (Alineamiento Temporal Dinámico o *Dynamic Time Warping* en inglés)[158]. Este algoritmo, originalmente concebido para el análisis y reconocimiento de la voz [156], permite encontrar el alineamiento óptimo entre dos series temporales de distinta longitud mediante técnicas no lineales de estiramiento o compresión de una de las series. El alineamiento permite obtener una medida de la distancia entre ambas, representando esta distancia el grado de similitud entre las series.

Con este planteamiento es factible comparar la similitud entre las diferentes secuencias de los movimientos de los músculos del rostro para diferentes estados afectivos, caracterizados éstos por el conjunto de expresiones faciales que los definen, esperando que las secuencias correspondientes a un mismo estado afectivo presenten mayores similitudes entre ellas y, por tanto, menores distancias que las obtenidas en comparación con secuencias de otros estados afectivos.

Para definir formalmente el funcionamiento de DTW, considérense dos series de valores X e Y correspondientes a dos secuencias temporales de valores con longitudes p y q, respectivamente:

$$X = x_1, x_2, \ldots, x_i, \ldots, x_p \ ; \ Y = y_1, y_2, \ldots, y_j, \ldots, Y_q \tag{3.10}$$

La función de alineamiento W para ambas series de valores se define como:

$$W = w_1, w_2, \ldots, w_s \ ; \ max(p, q) \leq s < p + q \tag{3.11}$$

donde $w_i = (j, k)$ es un par de elementos a comparar, correspondiendo j y k a índices en la series X e Y, respectivamente.

Para cada par de elementos w_k se obtiene una función de coste como una medida de distancia entre ambos:

$$d(w_i) = \delta(w_{ij}, w_{ik}) \tag{3.12}$$

Esta distancia proporcionará el grado de similitud entre los elementos comparados. La función δ habitualmente empleada es la distancia euclídea, aunque cualquier otra función podría ser usada para su cálculo, como por ejemplo la distancia de *manhattan*.

La función de alineamiento D óptima entre ambas series es aquella que minimiza la función D, es decir su distancia.

$$D(W) = \sum_{i=1}^{s} d(w_i) \tag{3.13}$$

3.5. Evaluación del rendimiento de los clasificadores

3.5.1. Validación cruzada

Una de las aproximaciones habitualmente empleadas en la evaluación del rendimiento de un sistema de aprendizaje consiste en la partición aleatoria del espacio total de muestras disponibles en dos conjuntos independientes: entrenamiento y test. El conjunto de muestras de entrenamiento se utiliza durante la fase de construcción del modelo de aprendizaje, mientras que el conjunto de test se emplea en la fase de evaluación del modelo.

El problema que plantea esta aproximación es que si el espacio total de muestras es relativamente pequeño, es posible que los conjuntos de entrenamiento y test no dispongan de suficientes muestras significativas para la construcción y evaluación del modelo. En estos casos, una alternativa es el empleo del método de validación cruzada (*cross-validation* en inglés). Este método divide aleatoriamente el espacio total de muestras N en K conjuntos con el mismo tamaño muestral, construyendo K modelos y empleando, en cada uno de los modelos, $K-1$ particiones como conjunto de entrenamiento, utilizando la partición restante para la validación del modelo (test).

Una variación al método de validación cruzada descrito anteriormente, consiste en la creación de N modelos, tomando como conjunto de entrenamiento en cada iteración $N-1$ muestras y dejando tan sólo una de ellas para la fase de test. A este procedimiento se le denomina validación cruzada dejando uno fuera (*leave-one-out cross-validation* en inglés).

El uso de técnicas de validación cruzada hace necesario promediar las medidas utilizadas en la evaluación del modelo en función del número de iteraciones o particiones realizadas.

3.5.2. Medidas de evaluación

Las herramientas habitualmente empleadas para medir la calidad predictiva de un sistema de aprendizaje supervisado son: las matrices de confusión junto con la valoración de medidas estadísticas como la exactitud, la precisión, la cobertura, el valor-F o las curvas ROC.

3.5.3. Matrices de confusión

Una herramienta que resulta de utilidad para la comprobación visual del rendimiento de un sistema, así como para el cálculo de las medidas de su calidad de predicción, es la matriz de confusión. Estas matrices se suelen representar en formato tabular, situando en las filas de la tabla las instancias reales de las clases (las observaciones) y en las columnas las predicciones obtenidas para cada clase, siendo las predicciones correctas aquellas que se encuentran en la diagonal de la matriz. Un ejemplo de una matriz de confusión para dos clases se muestra en la tabla 3.1, en la que las predicciones correctas se han marcado en color verde y las incorrectas en rojo.

Observaciones	**Predicciones**	
	Clase A	**Clase B**
Clase A	114	5
Clase B	4	108

Tabla 3.1: Ejemplo de una matriz de confusión para la clasificación de dos posibles clases

Tomando el ejemplo de la matriz de confusión mostrada en la tabla anterior, considerando que la clase A es equivalente a una categoría positiva y que la clase B lo es a una negativa, los términos verdadero positivo (VP) y verdadero negativo (VN) se refieren a las predicciones correctas para las clases A y B, respectivamente, el término falso positivo (FP) se define como el resultado de la predicción de una observación de Clase B como Clase A, mientras que un falso negativo (FN) se define como la predicción de una observación de Clase A como Clase B. En la tabla 3.2 se muestra la relación entre estas variables.

Observaciones	**Predicciones**	
	Clase A	**Clase B**
Clase A	VP	FN
Clase B	FP	VN

Tabla 3.2: Distribución de aciertos y fallos para la clasificaicón de dos posibles clases

3.5.4. Exactitud

La exactitud (*accuracy* en inglés) de un sistema binario con N muestras totales, se define como la proporción de clasificaciones realizadas correctamente frente al total.

$$Exactitud = \frac{VP + VN}{N} = \frac{predicciones\ correctas\ para\ cada\ clase}{total\ muestras} \tag{3.14}$$

Aunque la exactitud es una medida comúnmente utilizada para medir el rendimiento de un sistema de aprendizaje automático supervisado, es necesario ser cauto con su interpretación debido a que, en algunas circunstancias, su valor puede no ser representativo. Tomemos en consideración una matriz de confusión como la que se muestra en la tabla 3.3. En ella se observa que la distribución de muestras entre ambas clases se encuentra sesgada hacia la clase A. En estos casos los clasificadores suelen optar por clasificar prácticamente todas las muestras como la clase dominante (la de mayor número de muestras). En este ejemplo, la exactitud es de un 98,55 %, aunque es evidente que se están clasificando incorrectamente las muestras para la clase B y, por lo tanto, el sistema es completamente inútil desde el punto de vista de la predicción. Circunstancias como la descrita hacen necesario definir medidas alternativas para la evaluación de la bondad del sistema de aprendizaje, como son la precisión y la cobertura.

Observaciones	**Predicciones**	
	Clase A	**Clase B**
Clase A	818	2
Clase B	10	0

Tabla 3.3: Ejemplo de una matriz de confusión con resultados sesgados hacia una clase

3.5.5. Precisión

La precisión se define como la proporción de predicciones realizadas correctamente. La precisión para la clase A se puede definir formalmente como

$$Precisión(A) = \frac{VP}{VP + FP} = \frac{predicciones\ correctas\ para\ A}{total\ de\ predicciones\ como\ A} \tag{3.15}$$

3.5.6. Cobertura

La cobertura (*recall* en ingles), también conocida como sensibilidad o exhaustividad, se define como el porcentaje de aciertos frente al total. La cobertura para la clase A se define como

$$Cobertura(A) = \frac{VP}{VP + FN} = \frac{predicciones\ correctas\ para\ A}{muestras\ de\ A} \tag{3.16}$$

3.5.7. Valor-F

Un clasificador ideal presentaría una precisión y una cobertura igual a 1. Esta situación es conocida como utilidad teórica, difícilmente alcanzable en entornos reales. Una medida que relaciona ambas variables es el valor-F (también conocido como medida-F). El valor-F para una clase A se define formalmente como

$$Valor\text{-}F(A) = \frac{2 \times Precisión(A) \times Cobertura(A)}{Precisión(A) + Cobertura(A)} \tag{3.17}$$

3.5.8. Curvas ROC

Las curvas ROC, también denominadas como característica operativa del receptor (*Receiver Operating Characteristic* en inglés) [173], consisten en una representación gráfica del equilibrio del sistema de clasificación binaria mediante la comparación de la relación existente entre la tasa de predicciones correctamente realizadas (sensibilidad) y la de falsos positivos (1-especificidad).

Una medida del rendimiento del sistema de aprendizaje la proporciona el área bajo la curva ROC. Esta medida representa un indicador de la calidad predictiva del sistema, con una mayor precisión cuanto más próxima se encuentre la curva del borde superior izquierdo de la gráfica (mayor tasa de aciertos y menores falsos positivos) y, por el contrario, más impreciso cuanto más próximo se encuentre a la diagonal de 45 grados (se reduce la relación de aciertos frente a los falsos positivos). En general, valores del área bajo la curva ROC por encima de 0,70 representan un buen rendimiento del sistema de predicción, aunque este umbral dependerá del contexto de aplicación del sistema evaluado. La figura 3.9 muestra las curvas para tres clasificadores diferentes con diferente calidad predictiva.

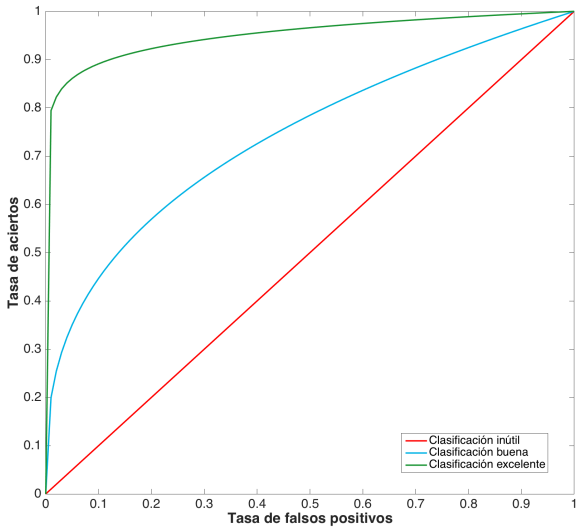


Figura 3.9: Ejemplo de curvas ROC para tres diferentes sistemas de aprendizaje

Parte II

Aportaciones en la detección emocional

En la última década la investigación en el área de la computación afectiva ha sido considerablemente activa, centrándose principalmente en la producción de aplicaciones que detectan y toman en consideración el estado afectivo del usuario como parte de un modelo de interacción. Los métodos que reconocen o predicen el estado afectivo del sujeto pueden usar diferentes fuentes de información, como vídeo, audio, dispositivos de entrada estándar, señales fisiológicas o información de profundidad.

Diversos son los enfoques habitualmente empleados en la detección del estado afectivo por computador. Entre ellos se encuentran los basados en técnicas de visión artificial, mediante aproximaciones holísticas o centrados en características locales; y las técnicas basadas en el estudio de señales fisiológicas, donde mediante estrategias de aprendizaje automático se intentan descubrir patrones que identifiquen los diferentes procesos emocionales del usuario. Las señales fisiológicas analizadas con más frecuencia son las correspondientes a EEG, ECG, conductividad de la piel, temperatura corporal, patrones respiratorios o EMG, entre otras.

En los capítulos que componen esta parte se introducen las aportaciones realizadas en el contexto de la detección emocional desde dos vertientes diferenciadas: visión por computador y el análisis de señales fisiológicas.

En el capítulo 4 se detalla el procedimiento llevado a cabo para el análisis y extracción de características faciales sobre los vídeos almacenados en la base de datos FEEDB (*Facial Expression and Emotion Database*) [175, 176], mediante el uso de dispositivos de bajo coste y poco intrusivos, como lo es el sensor Kinect de Microsoft. En el capítulo 5 se expone el análisis y la detección emocional realizada mediante la implementación de un enfoque holístico basado en la apariencia de la cara sobre un conjunto de imágenes estáticas contenidas en la base de datos CK+ (*The Extended Cohn-Kanade Dataset*) [115], fundamentado en la técnica de *Eigenfaces* [182], denominado en esta tesis como Eigenexpressions [120].

En el capítulo 6 se describe el análisis llevado a cabo sobre el conjunto de señales EEG contenidas en la base de datos MAHNOB-HCI [169] para la predicción de las emociones subyacentes, con el objeto de evaluar experimentalmente las posibilidades que ofrece este tipo de información en la detección del estado emocional del usuario.

Capítulo 4

Extensión de una base de datos de vídeos: FEEDB

Resumen
En este capítulo se describirá la primera de las aportaciones realizadas en la detección emocional mediante visión por computador. En él se explorarán las posibilidades que pueden ofrecer los dispositivos de bajo coste y poco intrusivos, como el sensor Kinect de Microsoft, en la extracción de información facial relacionada con el estado afectivo del usuario mediante la obtención de información adicional como las Unidades de Animación faciales, la posición y los ángulos de rotación de la cabeza, entre otras, sobre la base de datos afectiva FEEDB.

Contenidos

4.1. Descripción de FEEDB 45
4.2. Extracción de datos 46
4.3. Aproximación propuesta para la detección de estados afectivos sobre FEEDB 53
4.4. Clasificación de las muestras y limitaciones 54
4.5. Extensión de FEEDB 59
4.6. Conclusiones 61

La cara, sus expresiones y las deformaciones de los elementos faciales constituyen una valiosa fuente de información para la detección del estado emocional de las personas. Las emociones se transmiten mediante la manifestación de cambios en los elementos que componen la cara, como el levantamiento de las cejas y la apertura de la boca para denotar una situación de sorpresa, aunque en ocasiones dichos cambios no resultan tan evidentes, sino que se manifiestan de forma más sutil

mediante micro-expresiones como, por ejemplo, a través de un ligero fruncimiento de las cejas para denotar un sentimiento de desaprobación o desagrado.

Un planteamiento natural para el reconocimiento del estado afectivo del usuario es el análisis de su rostro mediante la extracción de características faciales que puedan sugerir su estado emocional, a través de la comparación de patrones y cambios en las estructuras faciales, de modo similar al proceso realizado por un observador humano cuando intenta juzgar el estado afectivo de su interlocutor.

Un requerimiento para el desarrollo de sistemas que sean capaces de detectar el estado afectivo del sujeto, es la disponibilidad de un corpus que pueda ser usado para entrenar y evaluar el rendimiento de los sistemas. Aunque existen bases de datos que integran diversas fuentes de información de carácter afectivo, no existe un corpus estándar que reúna todas las necesidades de un desarrollador de sistemas afectivos.

En este sentido, se consideró el empleo de una base de datos que contuviera suficientes secuencias de vídeo con sus correspondientes expresiones faciales etiquetadas con su estado emocional y afectivo subyacente, que aportara la posibilidad de obtener información adicional como la distancia del sujeto a la cámara (profundidad), información angular de la cabeza, posibilidad de extraer unidades de acción facial o equivalentes (AU), la geometría de la cara, etc. Entre varias alternativas evaluadas, se escogió FEEDB (*Facial Expression and Emotion Database*) [175, 176] como base de datos multimodal que proporcionaba todas las características anteriormente enumeradas.

Con el propósito de evaluar alternativas que se aproximen hacia el objetivo propuesto en esta tesis acerca de la evaluación del estado afectivo del usuario y su aplicación en entornos educativos interactivos, con la intención de explorar las posibilidades que pueden ofrecer los métodos y dispositivos de bajo coste y poco intrusivos para el estudiante, en este capítulo se describirá la primera de las aportaciones realizadas mediante técnicas de visión por computador. Con este fin, como dispositivo de bajo coste y reducida intrusividad, se empleó un sensor Microsoft Kinect para la extracción de información relacionada con el estado afectivo del usuario sobre los archivos de vídeo proporcionados en la base de datos afectiva FEEDB, con el objetivo de construir un conjunto de datos que pudieran resultar relevantes para la predicción del estado afectivo del usuario. Este corpus distribuye sus grabaciones en ficheros almacenados en formato XED, cuya especificación fue definida por Microsoft para los archivos grabados mediante Kinect, ofreciendo la posibilidad de ser tratados con un conjunto de herramientas ofrecidas gratuitamente por Microsoft, un sensor Kinect y su SDK (kit de desarrollo o *Software Development Kit* en inglés).

Para facilitar investigaciones posteriores, la información adicional extraída de los ficheros XED almacenados en FEEDB ha sido distribuida como extensión a la base de datos en formato texto, con el propósito de que estos nuevos datos puedan ser procesados de forma sencilla y con independencia de la plataforma por cualquier investigador con interés en el reconocimiento de expresiones faciales y de sus correspondientes emociones.

Antes de proceder con el detalle del procesado de los vídeos y extracción de los datos, se describirán las características de la base de datos empleada (FEEDB)

para la extracción y construcción del conjunto de datos que pudieran resultar relevantes para un sistema de predicción del estado afectivo de un usuario.

4.1. Descripción de FEEDB

FEEDB se encuentra disponible en dos posibles versiones. La primera versión [175] contiene 1650 grabaciones de 50 personas posando con 33 diferentes expresiones faciales. Estas grabaciones fueron recogidas íntegramente mediante un sensor Microsoft Kinect y se ofrece a la comunidad investigadora como un repositorio abierto. Cada grabación está compuesta por varios canales independientes y sincronizados de color y profundidad. Las grabaciones se proporcionan en el formato propietario entregado por Kinect: ficheros binarios con extensión XED. Un ejemplo de una secuencia de vídeo y algunas de las expresiones recogidas en la base de datos se muestra en la figura 4.1.

Figura 4.1: Ejemplo de una captura de un vídeo contenido en la primera versión de FEEDB y algunas de sus expresiones faciales

En la segunda versión de FEEDB [176], el conjunto de expresiones y el formato de las grabaciones fueron modificados. Esta nueva base de datos contiene 1550 grabaciones de 50 personas posando con 31 expresiones faciales. Adicionalmente, un conjunto de 10 emociones fueron expresadas de forma espontánea por 25 participantes, dando un total de 250 grabaciones adicionales. En esta segunda versión, el vídeo y la información sobre la profundidad fueron almacenados en formato AVI junto con algunos metadatos relacionados con la grabación.

En la tabla 4.1 se resumen las características de ambas versiones.

	FEEDB	
	Versión 1	**Versión 2**
Participantes	50	50
Emociones	33	31
Grabaciones	1650	1800 (1550+250)
Formato	XED	AVI

Tabla 4.1: Resumen de características de las dos versiones de FEEDB

Para las experimentaciones llevadas a cabo en el ámbito de esta tesis se decidió utilizar exclusivamente la primera versión de FEEDB y procesar los ficheros XED.

Entre los motivos que condujeron a la elección de esta versión y formato de datos se encuentran: 1) Los archivos XED proporcionan más información para el proceso de reconocimiento de la expresión facial que los vídeos en formato AVI facilitados en la segunda versión de la base de datos; 2) mediante el uso del SDK de Kinect para Windows es posible recuperar toda la información que el sensor capturó durante la grabación de las secuencias de vídeos, en concreto un total de 100 características faciales por fotograma, la posición de la cabeza del sujeto, sus ángulos de rotación, distancia del sujeto al sensor, además de 6 AU (*Animation Units*) y 11 SU (*Shape Units*) basadas el modelo Candide-3 [2] implementado por Kinect. Las AU son unidades de animación asociadas a diferentes elementos de la cara, mientras que las SU estiman la posición, forma y dimensiones de ciertas partes del rostro del usuario, como son la boca, las cejas, ojos, etc. Todos estos datos representan, *a priori*, una gran cantidad de información disponible para el análisis del estado afectivo de un usuario.

4.2. Extracción de datos

Los archivos XED incluidos en FEEDB pueden ser fácilmente procesados usando el software *Microsoft Kinect Studio* (MKS)[1], junto con el kit de desarrollo *Kinect for Windows Software Development Kit* (SDK)[2]. MKS puede usarse para remplazar el sensor Kinect como dispositivo de entrada, utilizando para ello una grabación en formato XED, mientras que el SDK para Kinect proporciona herramientas para la detección y seguimiento de la cara, así como múltiple información sobre sus características faciales. Sin embargo, el kit de desarrollo requiere un PC con Microsoft Windows instalado, junto con un sensor Kinect conectado que funcione, en este caso, como una licencia de uso.

Adicionalmente, Microsoft distribuye de forma gratuita otro SDK denominado *Microsoft Face Tracking SDK*[3] para el seguimiento de la cara y sus características faciales, que permite la creación de software de seguimiento facial en tiempo real. Este SDK es capaz de realizar el seguimiento de 100 puntos en un plano 2D, algunos de los cuales se muestran en la figura 4.2. La limitación de este kit de desarrollo radica en que únicamente puede ser usado en presencia de un sensor Kinect.

El SDK para el seguimiento facial devuelve la posición de la cabeza del usuario usando un sistema de coordenadas con sentido basado en la regla de la mano derecha con el origen en el sensor Kinect, el eje Z apuntando hacia el usuario y el eje Y hacia arriba. Proporciona, además, información angular de la posición de la cabeza en un espacio 3D, en concreto información sobre *pitch*, *roll* y *yaw*, con valores en el rango [-90, 90] grados. En la figura 4.3 se ilustra la dinámica de estos movimientos de cabeza.

[1] https://msdn.microsoft.com/en-us/library/dn785306.aspx
[2] https://developer.microsoft.com/es-es/windows/kinect
[3] https://msdn.microsoft.com/en-us/library/jj130970.aspx

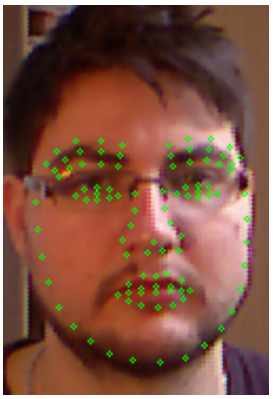


Figura 4.2: Algunos de los puntos en el plano 2D proporcionados por *Microsoft Face Tracking SDK*

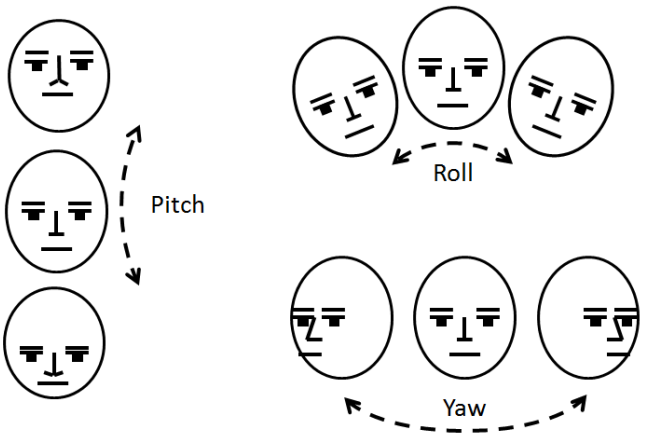


Figura 4.3: Información angular de la posición de la cabeza para *pitch*, *roll* y *yaw (fuente msdn.microsoft.com)*

Junto a la posición angular de la cabeza, se facilita información de los valores numéricos de 6 AU y de 11 SU según el modelo facial descrito en Candide-3 [2]. Este modelo consta de una máscara parametrizada de una cara humana estándar, con una descripción de su geometría en forma de polígonos (aproximadamente 100) controlados por AU (*Action Units*) globales y locales. Las AU globales corresponden a rotaciones de estos polígonos sobre el eje de un espacio tridimensional, mientras que las locales controlan sus deformaciones.

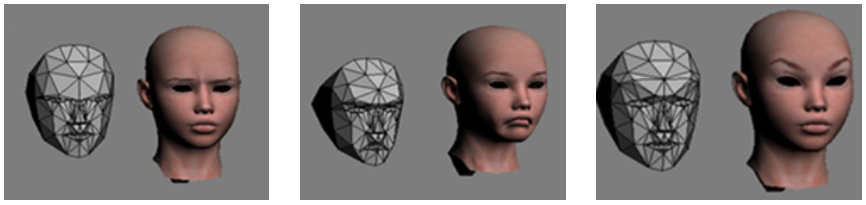
Las AU del modelo de Kinect son unidades de animación asociadas a diferentes elementos de la cara. Son deltas calculados desde una cara neutral expresados como valores numéricos en el rango [-1, 1]. En la figura 4.4 se muestran las seis AU obtenidas por Kinect y en la tabla 4.2 la interpretación de sus valores.

Las SU estiman la posición, forma y dimensiones de ciertas partes de la cara

(a) *AU0 - Levantamiento de la parte superior del labio (AU10 en Candide-3)*

(b) *AU1 - Apertura de la mandíbula (AU26/27 en Candide-3)*

(c) *AU2 - Fruncimiento de labios (AU20 en Candide-3)*



(d) *AU3 - Ceñimiento de las cejas (AU4 en Candide-3)*

(e) *AU4 - Caída de los extremos de la boca (AU13/15 en Candide-3)*

(f) *AU5 - Levantamiento de la parte exterior de las cejas (AU2 en Candide-3)*

Figura 4.4: Modelo de AU detectadas por el *Face Tracking SDK* de Kinect y su equivalencia con el modelo Candide-3 *(fuente msdn.microsoft.com)*

del usuario en referencia a una cara estándar, mediante la definición de sus vértices $\{x, y, z\}$. En la tabla 4.3 se relacionan las once SU entregadas por Kinect y se muestra su equivalencia con las definidas en el modelo Candide-3.

Con las herramientas ofrecidas por Microsoft y sus SDK se procesó cada uno de los vídeos en formato XED de FEEDB, extrayéndose información adicional que fue almacenada en formato texto para que pudiera ser procesada posteriormente de forma sencilla y con independencia de la plataforma. Este procesado se realizó en dos etapas, las cuales se ilustran en la figura 4.5.

En la primera etapa se hizo uso del *Microsoft Face Tracking SDK* para el desarrollo en Visual C++ de una herramienta de línea de comandos que procesaba en tiempo real las grabaciones XED y las transformaba en un fichero binario con la información original de color y de profundidad para cada fotograma, junto con la información de seguimiento facial descrita anteriormente. Esta herramienta se diseñó con el objetivo de eliminar la necesidad del uso de un sensor Kinect en el procesado posterior de las grabaciones incluidas en la base de datos. No obstante, para el procesamiento de los ficheros XED en esta etapa, se necesitó disponer de un sensor Kinect conectado a un equipo Microsoft Windows para la reproducción de las grabaciones, a la vez que la herramienta de línea de comandos se iniciaba para la conversión de los datos. La razón por la que se decidió usar un

ID	Descripción de la AU	Interpretación de los valores
AU0	Levantamiento de la parte superior del labio	0: Neutral (cubriendo dientes) 1: Completamente levantado -1: Completamente bajado
AU1	Apertura de la mandíbula	0: Cerrada 1: Completamente abierta -1: Cerrada (como 0)
AU2	Fruncimiento de labios	0: Neutral 1: Completamente estirados -0.5: Redondeados -1: Completamente redondeados (beso)
AU3	Ceñimiento de las cejas	0: Neutral -1: Completamente elevadas 1: Completamente bajadas
AU4	Caída de los extremos de la boca	0: Neutral -1: Sonrisa muy feliz 1: Muestra de tristeza
AU5	Levantamiento de la parte exterior de las cejas	0: Neutral -1: Completamente bajados (tristeza) 1: Elevados (sorpresa)

Tabla 4.2: Interpretación de los valores de las AU reportadas por el *Face Tracking SDK* de Kinect

formato binario en la conversión de las grabaciones fue para reducir el impacto en el rendimiento y los requerimientos de procesado. Aún así esta operación fue costosa en términos de almacenamiento, lo que requirió el uso de un disco SSD (de estado sólido o *Solid State Disk* en inglés) para poder leer y grabar los datos en tiempo real a una velocidad de 30 fotogramas por segundo.

En la segunda etapa los ficheros resultantes fueron procesados con una herramienta gráfica desarrollada en Java capaz de leer los binarios producidos por la herramienta de línea de comandos de la etapa anterior, creando estructuras de datos en formato texto apropiadas a su contenido, sin el requerimiento de disponer de un sensor Kinect ni de ningún otro hardware adicional. Este software fue específicamente desarrollado para dar soporte a la visualización y conversión de los datos contenidos en las grabaciones en formato XED de FEEDB. En la figura 4.6 se muestra una captura de pantalla del programa.

ID	Descripción de la SU	ID en Candide-3
SU0	Altura de la cabeza	0
SU1	Posición vertical de las cejas	1
SU2	Posición vertical de los ojos	2
SU3	Anchura de los ojos	3
SU4	Altura de los ojos	4
SU5	Separación de los ojos	5
SU6	Posición vertical de la nariz	8
SU7	Posición vertical de la boca	10
SU8	Anchura de la boca	11
SU9	Distancia vertical entre ambos ojos	-
SU10	Anchura de la barbilla	-

Tabla 4.3: Enumeración y descripción de las SU detectadas por el *Face Tracking SDK* de Kinect y su equivalencia con las SU definidas en el modelo Candide-3

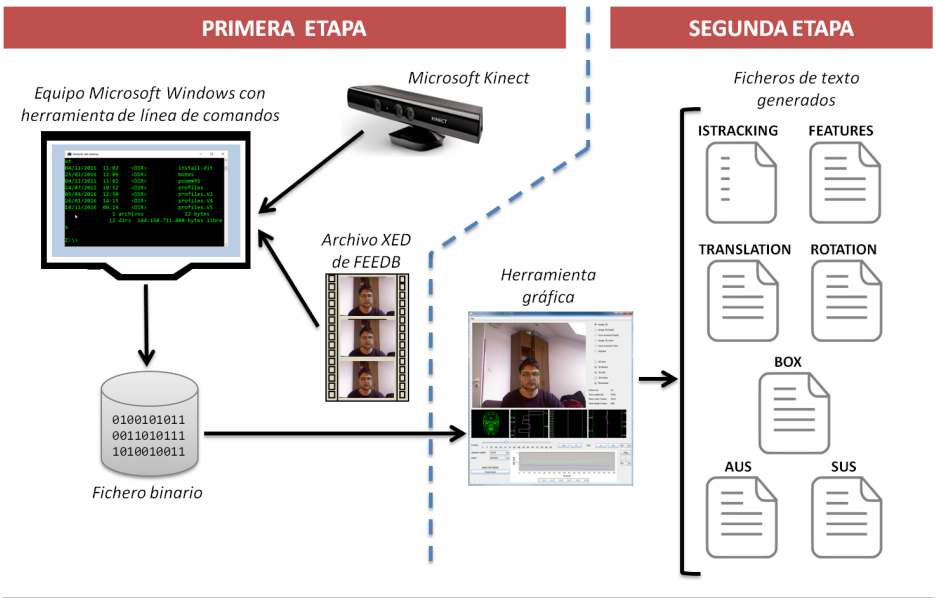


Figura 4.5: Proceso de la extracción de datos de los vídeos XED de FEEDB en dos etapas

La herramienta gráfica proporciona facilidades para la exportación de los datos a ficheros de texto independientes, conteniendo cada uno de ellos información sobre los ángulos de rotación de la cabeza (etiquetado en la imagen como ROT), de translación (TRA), AUs, SUs, etc., para cada uno de los fotogramas de la grabación de vídeo original, y en particular:

- ISTRACKING: Reporta información sobre si el seguimiento de la cara está

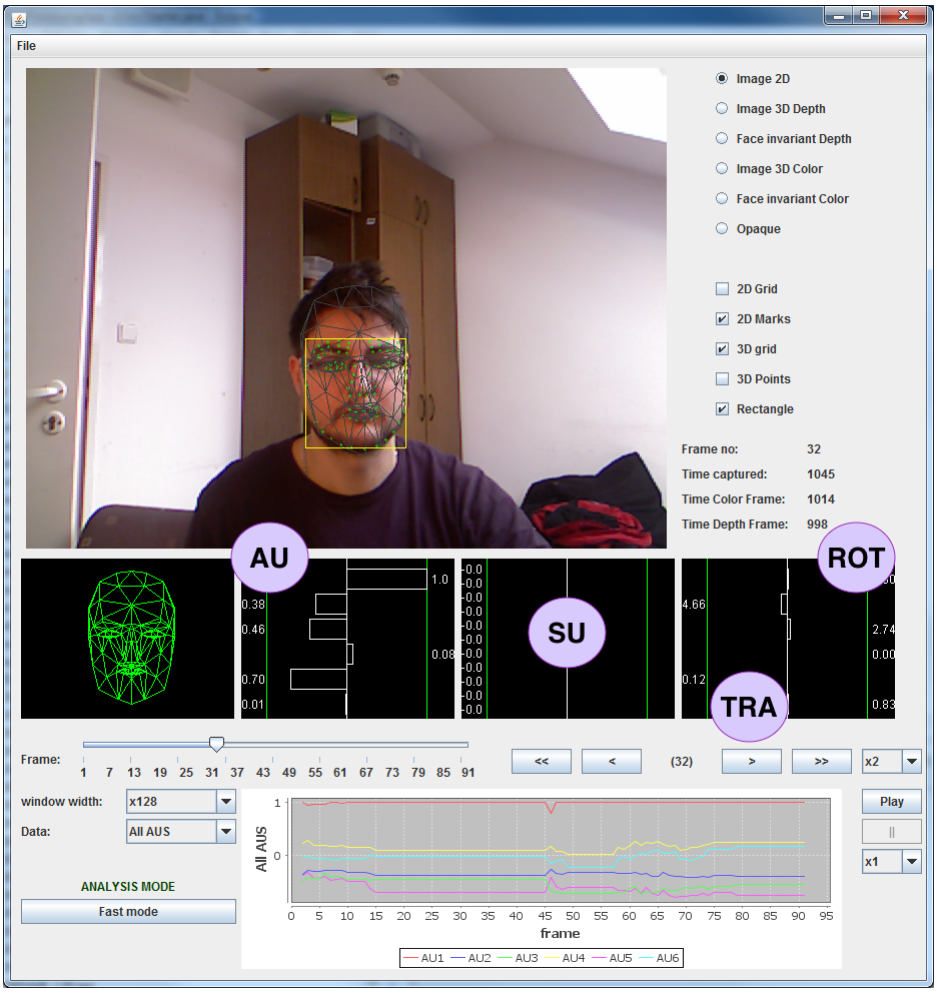


Figura 4.6: Herramienta gráfica usada en la visualización y conversión de los datos binarios obtenidos de las grabaciones originales en formato XED de FEEDB

activo (valor 1) o si, por el contrario, el sensor no ha sido capaz de realizarlo (valor 0), en cuyo caso el resto de la información deja de ser válida al no existir seguimiento facial.

- TRANSLATION. Indica la posición de la cabeza del usuario en el espacio de la cámara, tal y como se define en la figura 4.7. Las unidades se encuentran expresadas en metros, tomándose como punto de referencia al sensor Kinect.

- ROTATION. Contiene información sobre los ángulos de rotación de la cabeza para *pitch*, *roll* y *yaw* (rotaciones sobre los ejes X, Z e Y, respectivamente), en este mismo orden.

- BOX. Las coordenadas X e Y del rectángulo que define la cara. En concreto las coordenadas de la esquina superior izquierda y de la esquina inferior derecha del rectángulo que contiene la cara detectada. Estos valores se proporcionan en píxeles según el espacio de vídeo definido en la figura 4.7.
- FEATURES. Las coordenadas X e Y en el espacio de vídeo para cada uno de los 100 puntos de seguimiento detectados en la cara.
- AU. Definen 6 valores en coma flotante en el rango [-1, 1], correspondientes a los niveles de activación de las 6 AU proporcionadas por el SDK.
- SU. Definen 11 valores en coma flotante correspondientes a la estimación de la geometría de la cara del usuario: altura de la cabeza, posición vertical de las cejas, posición vertical de los ojos, anchura, altura y separación de los ojos, posición vertical de la nariz, posición vertical de la boca, anchura de la boca, distancia vertical entre ambos ojos y anchura de la barbilla.

Figura 4.7: Espacio de la cámara del sensor Kinect medido en metros y representación en un espacio de vídeo estándar de 640x480 píxeles de resolución

Todos los ficheros de texto producidos por la herramienta gráfica siguen la misma estructura: cada una de sus filas representa un fotograma; y la información contenida en cada fila se encuentra separada por espacios. Esta estructura permite fácilmente la carga y el tratamiento de los datos en otras aplicaciones como Mathworks MATLAB[4], Microsoft Excel o cualquier otra aplicación estándar de análisis de datos. Por ejemplo, en Excel los datos pueden ser cargados en una hoja mediante la utilidad de importación de ficheros de texto que el programa proporciona. En el caso específico de MATLAB los ficheros pueden ser cargados mediante las utilidades de importación que incorpora la propia herramienta. En la figura 4.8 se muestra un ejemplo de un gráfico generado en MATLAB para la información sobre las 6 AU y para el desplazamiento de la cabeza a lo largo de los ejes X, Y y Z.

El número de elementos de información obtenidos para cada fotograma y para cada uno de los ficheros se encuentra resumido en la tabla 4.4

[4]http://es.mathworks.com/products/matlab

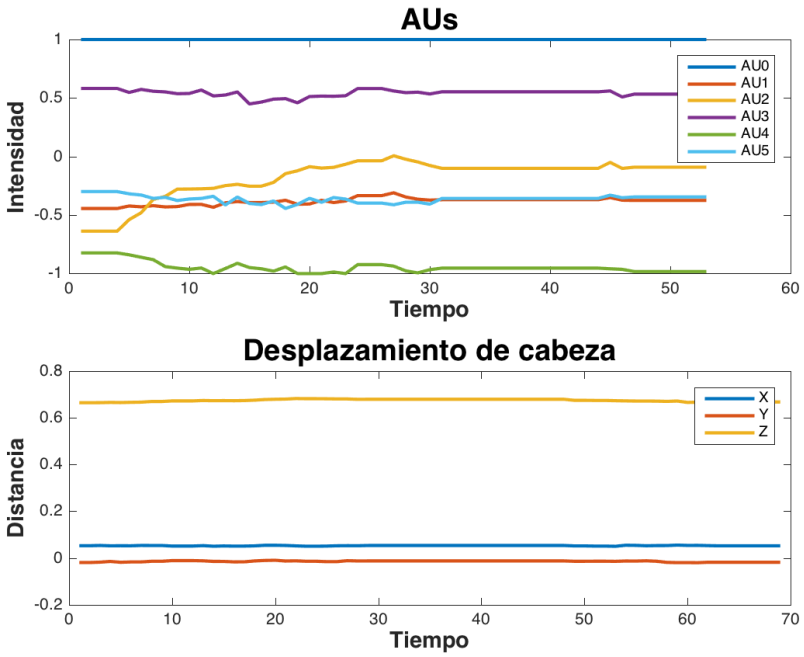


Figura 4.8: Visualización de los datos de FEEDB extraídos con Kinect en Mathworks MATLAB

Sufijo del fichero	Elementos de información por fotograma
ISTRACKING	1
TRANSLATION	6
ROTATION	3
BOX	4
FEATURES	200
AU	6
SU	11

Tabla 4.4: Lista de ficheros y elementos de información generados para cada grabación XED de FEEDB

4.3. Aproximación propuesta para la detección de estados afectivos sobre FEEDB

Del conjunto de estados afectivos incluidos en FEEDB, se consideraron aquellos que podían resultar relevantes en un entorno educativo, en concreto: Sorpresa (neutral), Sorpresa (moderadamente positiva), Sorpresa (muy positiva), Sorpresa (negativa), Sonrisa (positiva), Sonrisa (débilmente positiva), Sonrisa (negativa),

Placer, Excitación, Tristeza, Bostezo, Duda, Aburrimiento, Frustración, Ira, Concentración y Atención.

Una vez seleccionado el conjunto de vídeos correspondientes a esta selección de estados afectivos, se propuso una aproximación basada en el análisis de la secuencia de movimientos de las AU entre fotogramas para cada individuo y expresión, como elementos diferenciadores entre las distintas expresiones faciales.

Un algoritmo habitualmente empleado en el análisis de secuencias de valores a lo largo de un eje temporal es DTW (alineamiento temporal dinámico o *Dynamic Time Warping* en inglés)[156, 158]. En la aproximación propuesta en este capítulo, se utilizará DTW como método para analizar la similitud entre diferentes secuencias de expresiones faciales, con el objetivo de clasificarlas atendiendo al criterio de la menor distancia entre ellas. Para ello se asumirá que dos expresiones faciales correspondientes a un mismo estado afectivo deberían presentar mayor similitud y, en consecuencia, una menor distancia entre ellas, que las correspondientes a secuencias de expresiones faciales de diferentes estados afectivos.

4.4. Clasificación de las muestras y limitaciones

A partir de la secuencia de movimientos de las seis AU extraídas para cada sujeto y emoción mediante la implementación software descrita en el apartado 4.2, se calcularon los desplazamientos con respecto al fotograma anterior para cada uno de los fotogramas de la secuencia completa de vídeo, con el objetivo de poder aplicar posteriormente el algoritmo DTW y obtener una matriz con el cómputo de las distancias de los desplazamientos entre sujetos y emociones.

Para el cálculo de las distancias se desarrolló una implementación estándar del algoritmo DTW, utilizando para el cómputo de las distancias entre muestras la distancia euclídea y obteniendo como resultado una matriz simétrica de distancias para cada par de la combinación sujeto–emoción.

Con la asunción de que el conjunto de expresiones faciales correspondientes a un estado afectivo concreto deberían tener una secuencia similar de deformaciones de los músculos faciales, caracterizadas éstas por los movimientos de sus AU, y que diferentes expresiones deberían presentar secuencias con poca similitud, calculada ésta a partir de la matriz de distancias obtenida mediante DTW, se intentó predecir la emoción de cada secuencia mediante la selección del estado afectivo que presentaba la menor distancia.

En la figura 4.9 y de forma individualizada por cada AU en la figura 4.10, se muestra la secuencia de movimientos de las seis AU para dos estados diferentes (placer y bostezo) y distintos sujetos, observándose que existen diferencias siginificativas en las formas de las señales que indican que las expresiones faciales son muy diferentes entre sí.

Por otro lado, en la figura 4.11 se muestra la secuencia de movimientos de las seis AU para dos sujetos diferentes y un mismo estado (placer), donde se puede observar que tampoco existe una alta similitud en la secuencia de movimientos entre ambas. Con un nivel mayor de detalle, en la figura 4.12 se comparan de forma individual cada una de las seis AU. Para este ejemplo concreto, se puede

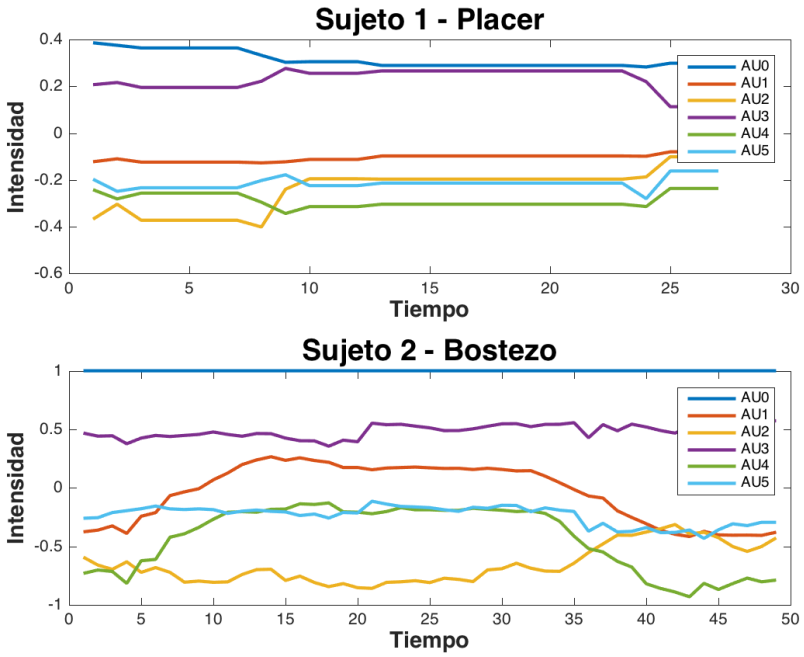
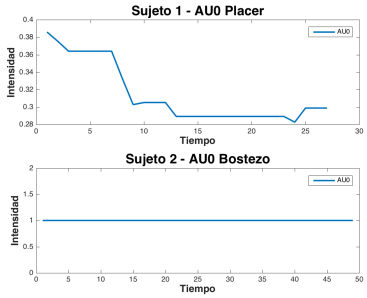


Figura 4.9: Comparativa de las secuencias de movimientos de las AU para dos estados diferentes (placer y bostezo) y distintos sujetos

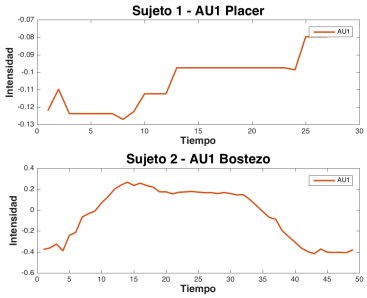
observar que mientras en algunas secuencias la señal presenta similitudes (AU3 y AU5), en otras presentan diferencias considerables (AU0, AU1, AU2 y AU4). Estas amplias diferencias en las señales, entre otras razones, hicieron que el sistema no pudiera determinar la expresión correcta en la mayoría de los casos.

Además de las amplias diferencias entre señales para un mismo estado, las razones por las que las muestras no pudieron ser clasificadas correctamente mediante la aproximación propuesta fueron varias. En primer lugar, por las limitaciones físicas que impone el sensor. Kinect es capaz de trabajar en dos modos: *default* y *near*. En el modo *default* se obtiene una mayor precisión en las mediciones para objetos situados a una distancia entre 80 centímetros y 4 metros de la cámara del sensor. En el modo *near* para objetos situados entre 40 centímetros y 3 metros. Debido a que las operaciones de seguimiento descansan sobre la información capturada por el sensor, éstas no funcionan para sujetos situados a distancias menores de 40 centímetros al encontrarse fuera de la distancia mínima soportada por Kinect. Por otro lado, el seguimiento de la cara únicamente se consigue cuando la inclinación de la cabeza del usuario se encuentra dentro de unos límites; en concreto por debajo de 20, 90 y 45 grados para *pitch*, *roll* y *yaw*, respectivamente, aunque el seguimiento es más preciso con valores por debajo de 10, 45 y 30 grados, respectivamente.

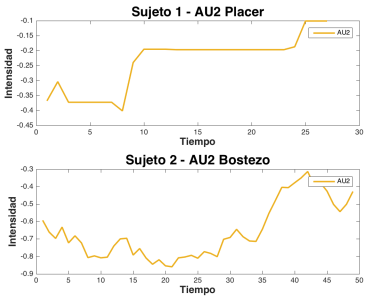
Debido a estas limitaciones, el conjunto de muestras de las que se pudo extraer



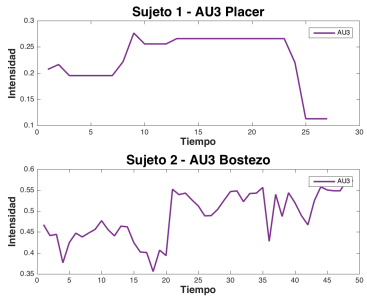
(a) *AU0*



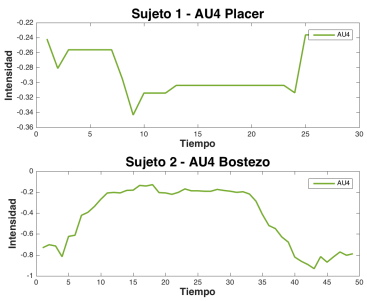
(b) *AU1*



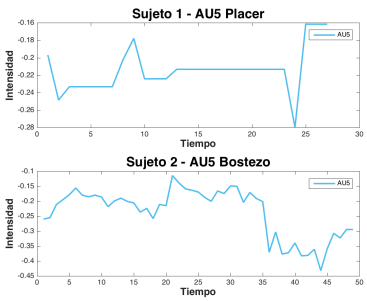
(c) *AU2*



(d) *AU3*



(e) *AU4*



(f) *AU5*

Figura 4.10: Comparativa de las secuencias individuales de movimientos de cada AU para dos estados diferentes (placer y bostezo) y distintos sujetos

información con Kinect fue relativamente reducido, principalmente porque gran parte de los sujetos grabados en FEEDB se situaron muy próximos al sensor, a una distancia cercana a los 40 centímetros, lo que provocó que Kinect no fuera capaz de realizar el seguimiento de la cabeza del usuario, dando lugar, incluso, a situaciones en las que el sistema alternaba constantemente entre fases de detección y no detección produciendo datos no válidos para su procesado. Esta limitación

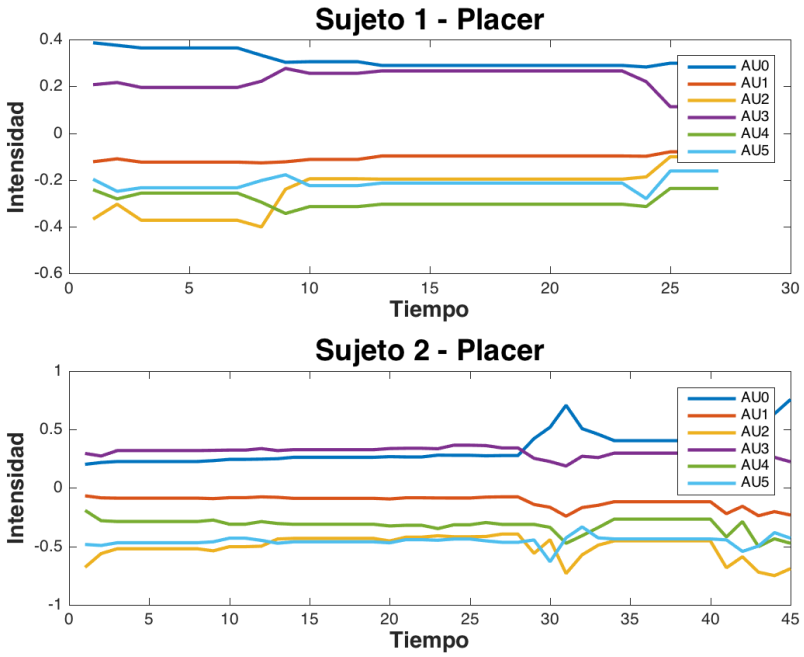
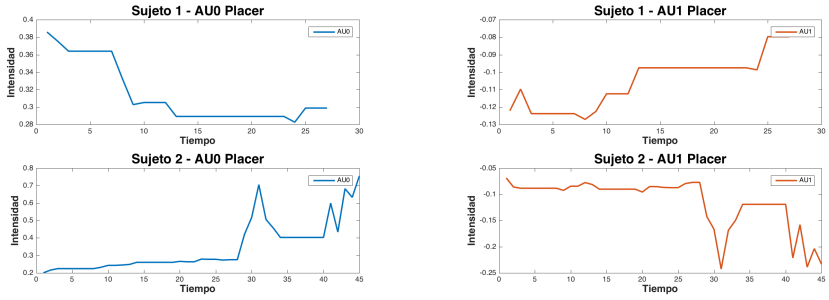


Figura 4.11: Comparativa de las secuencias de movimientos de las AU para un mismo estado (placer) y distintos sujetos

supuso que para el subconjunto de 17 expresiones previamente seleccionadas únicamente se pudiera procesar la mayoría de las emociones para tan solo 5 participantes (entre 10 o más expresiones). En el resto de los casos únicamente se pudo extraer información de 1 a 7 expresiones por individuo, con una media de cinco muestras por expresión. Además, las secuencias que recogen la expresión son muy cortas, principalmente por el tiempo que tarda Kinect en detectar el rostro y, en consecuencia, en proporcionar información sobre las AU, información, por otro lado, poco precisa por la imposibilidad de obtener información sobre las SU que componen la cara.

Por otro lado, debido a que la mayoría de las grabaciones fueron capturadas con sus usuarios situados enfrente de la cámara, los valores registrados para los movimientos de *pitch*, *roll* y *yaw* ofrecieron escasa información para el reconocimiento del estado afectivo del sujeto.

Por último, el *Face Tracking SDK* también es capaz de calcular un conjunto de 11 SU como parte del modelo Candide-3. Estas SU son usadas para adaptar la forma de la cara detectada a un modelo estándar de cara en 3D. Cada SU está relacionada con un aspecto particular de la cara humana, como la anchura y altura de los ojos, la posición vertical de la nariz y de la boca o la distancia entre los ojos. El problema encontrado al analizar las secuencias de vídeo de FEEDB es que debido a que el sensor tarda aproximadamente 2 minutos en converger, es



(a) *AU0* **(b)** *AU1*

(c) *AU2* **(d)** *AU3*

(e) *AU4* **(f)** *AU5*

Figura 4.12: Comparativa de las secuencias individuales de movimientos de cada AU para un mismo estado (placer) y distintos sujetos

decir en detectar y ajustar las SU para un usuario dado, no fue posible obtener esta información de los vídeos contenidos en la base de datos.

4.5. Extensión de FEEDB

Tras evaluar la información contenida en FEEDB y realizar la extracción de los datos a partir de ficheros XED con Kinect, se decidió que los datos obtenidos en el ámbito de esta tesis podrían resultar de utilidad a otros investigadores, por lo que se llegó a un acuerdo con el responsable del corpus para incluirlos en una nueva versión de FEEDB, realizando, de este modo, una extensión de la base de datos con la inclusión de nuevos datos en un formato fácilmente procesable para cada una de las grabaciones.

La principal motivación para la extensión de FEEDB fue el formato de las grabaciones. La primera versión usa el formato propietario XED, mientras que la segunda utiliza el formato AVI. Aunque este último es un formato accesible, el procesamiento de los fotogramas de vídeo para él análisis de la cara es costoso, así como la dificultad para encontrar herramientas robustas para su análisis, como por ejemplo librerías de reconocimiento facial. Por otro lado, los archivos XED proporcionan mayor información sobre el rostro y, por ende, sobre la expresión facial a analizar.

Aún así aunque los archivos XED pueden ser fácilmente procesados usando las herramientas de desarrollo software para Kinect proporcionadas por Microsoft, su SDK requiere un PC con Microsoft Windows instalado junto con un sensor Kinect conectado que funcione como una licencia de uso.

Con el propósito de evitar la necesidad de usar un sensor Kinect en cualquier investigación o proyecto futuro que desee trabajar con FEEDB, se decidió facilitar los datos obtenidos a partir del procesado de los archivos XED en un formato más simple (formato texto), extendiendo de este modo las posibilidades ofrecidas en la primera versión de FEEDB [119].

Todos los ficheros de texto producidos por la herramienta gráfica (*ISTRACKING, FEATURES, TRANSLATION, ROTATION, BOX, AUX y SUS*) siguen la misma estructura: cada una de sus filas representa un fotograma; y la información contenida para el fotograma se encuentra definida con valores en coma flotante separados, cada uno de ellos, por espacios. De este modo, todos los ficheros generados para una secuencia de vídeo compuesta por 300 fotogramas correspondientes a 10 segundos de grabación, a razón de 30 fotogramas por segundo, contendrán 300 filas con la siguiente información.

- Fichero *_ISTRACKING_.TXT*: Este fichero únicamente contendrá una columna con dos posibles valores $\{1, 0\}$ para cada una de las filas correspondientes a cada fotogramas de la secuencia de vídeo, con información sobre la validez del seguimiento de la cara. Un valor 1 indica que el seguimiento se encuentra activo para el fotograma, mientras que un valor 0 indica que el seguimiento no ha sido posible realizarlo, en cuyo caso la información contenida en el resto de ficheros generados, para ese fotograma concreto, dejará de ser válida al no existir seguimiento facial. Es importante destacar que en los casos en los que el sensor alterna entre fases de detección y no detección, aparecerán, alternativamente, filas con valores 1 y 0, validando o

invalidando los datos de sus correspondientes fotogramas en el resto de los archivos generados.

- Fichero *_BOX_.TXT*: Define las coordenadas X e Y en píxeles del rectángulo que define la cara detectada por Kinect. Contiene dos pares de valores: un par para la definición de las coordenadas correspondientes a la esquina superior izquierda del rectángulo que contiene el rostro; y un segundo par para las coordenadas de la esquina inferior derecha del mismo.

- Fichero *_TRANSLATION_.TXT*: Contiene las coordenadas X, Y y Z de la cabeza del usuario en el espacio de la cámara. Las unidades se encuentran expresadas en metros, tomándose como punto de referencia al sensor Kinect.

- Fichero *_ROTATION_.TXT*: En este archivo se almacena información angular de la cabeza para los movimientos de *pitch*, *roll* y *yaw*, representando los ángulos de rotación sobre los ejes X, Z e Y, respectivamente. Cada fila contiene una terna de valores en el rango [-90, 90] con la expresión del ángulo de inclinación para cada uno de las tres posibles rotaciones de cabeza.

- Fichero *_FEATURES_.TXT*: En cada fila se proporcionan 100 pares de coordenadas X e Y sobre el espacio de vídeo de Kinect para cada uno de los puntos de seguimiento detectados en la cara.

- Fichero *_AUS_.TXT*: Para cada fotograma se facilitan 6 valores en el rango [-1, 1] correspondientes a los niveles de activación de las 6 AU proporcionadas por el SDK, según la interpretación de los valores reportados por el *Face Tracking SDK* de Kinect descritos en la tabla 4.2.

- Fichero *_SUS_.TXT*: En este archivo se definen los 11 valores correspondientes a la estimación de la geometría de la cara del usuario: altura de la cabeza, posición vertical de las cejas, posición vertical de los ojos, anchura, altura y separación de los ojos, posición vertical de la nariz, posición vertical de la boca, anchura de la boca, distancia vertical entre ambos ojos y anchura de la barbilla. Debido al tiempo que el sensor Kinect tarda en obtener esta información (aproximadamente 2 minutos) y la escasa duración de los vídeos contenidos en FEEDB (inferior a 2 minutos), no fue posible obtener información de los SU, motivo por el que la mayoría de estos archivos no contienen información útil.

Con esta extensión se pretendió obtener dos beneficios: en primer lugar es posible procesar la información contenida originalmente en ficheros XED con cualquier tipo de plataforma software y hardware; en segundo lugar es posible emplear sobre estos nuevos datos cualquier técnica de clasificación existente, focalizando los esfuerzos únicamente en la extracción de conocimiento sin la necesidad de tener que trabajar con formatos o dispositivos propietarios y, en ocasiones, con limitada accesibilidad.

La extensión de la base de datos consistió en la producción de información adicional para una combinación total de 88 sujetos y estados afectivos. Esta

información, detallada en la tabla 4.5, es la que fue propuesta para su inclusión en una versión extendida de FEEDB con el propósito de proporcionar datos adicionales que puedan ser usados para evaluar la compleja y esquiva naturaleza del estado afectivo del usuario.

Sujeto	01. Sorpresa (neutral)	01b. Sorpresa (moderadamente positiva)	01c. Sorpresa (muy positiva)	01d. Sorpresa (negativa)	02. Sonrisa (positiva)	02a. Sonrisa (débilmente positiva)	02c. Sonrisa (negativa)	04. Placer	05. Excitación	07. Tristeza	13. Bostezo	14. Duda	15. Aburrimiento	17. Frustración	18. Ira	19. Concentración	21. Atención	**NÚMERO DE MUESTRAS**
490	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	17
378		✓	✓	✓	✓				✓						✓			6
518	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		16
669	✓		✓				✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	12
388									✓	✓	✓	✓		✓	✓	✓		7
363																✓	✓	2
478			✓	✓	✓			✓	✓		✓	✓	✓	✓	✓	✓		11
317					✓													1
290	✓	✓		✓	✓					✓	✓	✓	✓	✓			✓	10
374		✓			✓							✓			✓		✓	5
473														✓				1
Muestras	4	5	5	5	7	2	3	4	6	5	6	7	4	7	7	6	5	88

Tabla 4.5: Grabaciones y estados afectivos / expresiones sobre las que se ha extendido la información existente en FEEDB

4.6. Conclusiones

La detección y evaluación del estado afectivo de una persona es una tarea compleja. Bases de datos como FEEDB proporcionan información valiosa a la comunidad de investigadores y constituyen una base experimental de utilidad para sistemas basados en el análisis de datos sobre emociones y estados afectivos, especialmente para aquellas implementaciones basadas en dispositivos de bajo coste y poco intrusivos, como es el caso del sensor Kinect.

Con la información obtenida a partir del procesado de los archivos XED originales y con la idea de que todas las secuencias correspondientes a un estado afectivo concreto deberían presentar similares sucesiones en las deformaciones de los músculos faciales, se construyó un conjunto de datos etiquetados con su correspondiente estado afectivo y con información sobre la expresión facial detectada. Esta información se concretó en un conjunto de ficheros de texto con datos sobre la posición de la cabeza, ángulos de rotación, características faciales y AU, principalmente.

Sobre este conjunto de datos se propuso una aproximación basada en el algoritmo DTW para el reconocimiento de expresiones a partir de la secuencia de movimientos de las AU entre fotogramas, utilizando como medida de similitud el cómputo de la distancia entre secuencias. La aproximación basada en DTW y el cálculo de la distancia entre secuencias no fue capaz de arrojar resultados concluyentes en la clasificación de expresiones faciales, principalmente por el escaso número de muestras por emoción obtenidas para la clasificación y su corta duración temporal. Entre los motivos que incidieron en la consecución del reducido conjunto de muestras se encuentran las limitaciones físicas impuestas por Kinect, en especial su dificultad para detectar rostros a distancias inferiores a los 40 centímetros, aspecto que hizo que muchas de las grabaciones contenidas en FEEDB no resultaran de utilidad para el análisis de expresiones faciales o que, incluso, aquellas que pudieron procesarse alternaran entre secuencias de detección y no detección del rostro. Estas circunstancias, unidas al tiempo que tarda Kinect en converger las SU y la breve duración de las grabaciones recogidas en el corpus, aspectos que, colateralmente, también afectaron al número de AU disponibles para la predicción de la emoción subyacente, no hicieron factible la posibilidad de obtener predicciones fiables sobre las secuencias de vídeo grabadas.

No obstante, se consideró que los datos extraídos y almacenados en un formato fácilmente legible y procesable, características que presenta el formato escogido (ficheros de texto), podrían resultar de utilidad a posibles investigadores interesados en la computación afectiva. Por este motivo, el conjunto de información obtenido en este trabajo a partir de los ficheros en formato XED almacenados en FEEDB se distribuyó como extensión al corpus con el propósito de que estos nuevos datos pudieran ser procesados de forma sencilla, con independencia de la plataforma y sin necesidad de disponer de un sensor Kinect, en cualquier sistema típico de clasificación, por ejemplo simplemente agrupando por su similitud las diferentes expresiones y estados afectivos. En la tabla 4.6 se describe el número de muestras y sujetos por categoría cuando las emociones se encuentran agrupadas según un criterio de similitud.

Otra estrategia de clasificación alternativa y ampliamente utilizada consiste en la distribución de las emociones y estados afectivos en las dimensiones de Valencia y Activación para cada una de las emociones consideradas. Esta aproximación se ha utilizado en [169]. Con este enfoque, las ocho emociones de la tabla 4.6 pueden ser asignadas en tres clases para cada dimensión, como se describe en las tablas 4.7 y 4.8.

El objetivo principal del trabajo fue la evaluación de un método de bajo coste y poco intrusivo como Kinect para la detección de expresiones faciales y

Emoción / Estado	**Estados FEEDB**	**Muestras**	**Sujetos**
Sorpresa	(01+01b+01c+01d)	19	7
Felicidad	(02+02a+02c+04)	16	8
Excitación	(05)	6	6
Tristeza	(07)	5	5
Aburrimiento	(13+15)	10	6
Frustración	(14+17)	14	8
Ira	(18)	7	7
Concentración	(19+21)	11	8
TOTAL	**88**	**55**	

Tabla 4.6: Emociones y estados afectivos de FEEDB agrupados en categorías

Clases de Valencia	**Emociones / Estados**	**Número de muestras**
Negativa	Tristeza, Aburrimiento, Frustración, Ira	36
Neutral	Sorpresa, Concentración	30
Positiva	Felicidad, Excitación	22
TOTAL		**88**

Tabla 4.7: Emociones y estados de FEEDB clasificados en clases de Valencia

Clases de Activación	**Emociones / Estados**	**Número de muestras**
Relajación	Tristeza, Aburrimiento	15
Media	Felicidad	16
Excitación	Sorpresa, Excitación, Frustración Ira, Concentración	57
TOTAL		**88**

Tabla 4.8: Emociones y estados de FEEDB clasificados en clases de Activación

la predicción del estado afectivo subyacente, siendo su principal aportación la extracción y recopilación de características faciales a partir de las grabaciones originales almacenadas en el formato propietario XED de Microsoft, en concreto de 100 características faciales, la posición de la cabeza del usuario, sus diferentes ángulos de inclinación y seis unidades de animación facial (AU). La información relativa a la geometría de la cara, caracterizada por las unidades de forma facial (SU), no pudieron obtenerse por el tiempo que tarda Kinect en converger y ajustar dichas unidades (aproximadamente 2 minutos), así como por la escasa duración de las grabaciones contenidas en FEEDB.

La información de extensión extraída fue almacenada en ficheros de texto independientes fácilmente procesables para una combinación total de 88 sujetos-emociones, con el objetivo de que pueda ser utilizada por cualquier investigador sin la complejidad de conocer la estructura de datos del formato XED, evitando

la necesidad de disponer de un sensor Kinect conectado al ordenador en el que se desee utilizar los datos proporcionados por FEEDB, con independencia de la plataforma hardware o software utilizada.

Por otro lado, la aproximación seguida en este proyecto podría servir de base a otros investigadores en la extracción de datos obtenidos con un sensor Microsoft Kinect, así como el conocimiento sobre las limitaciones detectadas en el estudio en lo referente a la distancia entre el sujeto y el sensor o el tiempo necesario en la grabación para que el sistema pueda converger y calcular las SU, pueden resultar de utilidad en futuros proyectos realizados con Kinect.

Capítulo 5

Detección emocional sobre imágenes estáticas: Eigenexpressions

Resumen
En este capítulo se describirá la segunda aportación realizada en el área de la detección emocional mediante técnicas de visión artificial. En este caso se adoptó un enfoque holístico y una aproximación específica desarrollada en el ámbito de esta tesis, denominada Eigenexpressions, para el reconocimiento de expresiones faciales sobre imágenes estáticas.

Contenidos

5.1. Introducción 66
5.2. Aproximación propuesta: Eigenexpressions 67
5.3. Configuración de Eigenexpressions 69
5.4. Evaluación de Eigenexpressions 72
5.5. Extensión de Eigenexpressions mediante máscaras de expresiones faciales 75
5.6. Evaluación de la extensión mediante máscaras 78
5.7. Conclusiones 81

En el capítulo anterior se empleó una aproximación basada en la identificación de determinadas características y regiones faciales específicas sobre las AU detectadas por Microsoft Kinect para la detección del estado afectivo del usuario. Un modo alternativo de abordar el problema de la detección de la expresión facial consiste en el empleo de un enfoque holístico basado en la apariencia de la cara.

En este capítulo se describe la implementación y evaluación de un sistema de aprendizaje supervisado en dos etapas, basado en la técnica holística de *Eigenfaces* [182] y denominado en este trabajo como Eigenexpressions [120].

5.1. Introducción

Muchas de las técnicas holísticas de clasificación basadas en la apariencia [21] han sido probadas con éxito en el área del reconocimiento facial, aunque el problema con el que se suelen encontrar estos métodos de reconocimiento de patrones que usan una representación de la cara mediante píxeles es la alta dimensionalidad de los datos. Esto es así porque en una simple imagen de 512 x 512 píxeles la dimensionalidad es de 262.144 unidades de información.

No obstante, en los problemas de reconocimiento facial la intrínseca dimensionalidad del espacio de la cara es mucho menor que la del espacio de la imagen debido a que las caras son similares en apariencia y contienen significantes regularidades estadísticas [49]. El método *Eigenfaces* [182] ha sido ampliamente utilizado para reducir la dimensionalidad del espacio de entrada de la cara. Este método se basa en el Análisis de Componentes Principales (PCA), realizando un análisis holístico de la cara y proyectándola sobre un espacio dimensional reducido sobre el que es realizado el reconocimiento facial.

Relacionado con el reconocimiento facial se encuentra el reconocimiento de expresiones en rostros humanos y la detección de la emoción subyacente. En este área, son muchos los métodos que basan el reconocimiento de la expresión en ciertas características geométricas de la cara que representan la forma, posición relativa o las deformaciones de los elementos representativos de ella, tales como la boca, la nariz o las cejas. En muchas ocasiones [56, 138, 109] esta información se utiliza para identificar las unidades de acción facial (AU) definidas en el sistema de codificación de acciones faciales de Ekman (Facial Action Coding System, FACS) [60], en el que se asocian ciertas combinaciones de AU con alguna de las seis emociones primarias: Alegría, Ira, Miedo, Asco, Sorpresa y Tristeza.

Aunque en general las aproximaciones basadas en la apariencia han recibido considerablemente una menor atención en el campo del reconocimiento de las expresiones que en el del reconocimiento facial, entre los diversos trabajos realizados en el ámbito de esta tesis, se desarrolló una aproximación holística basada en la apariencia para el reconocimiento de expresiones faciales sobre imágenes estáticas. La solución se fundamentó en el método estándar de *Eigenfaces* para el reconocimiento facial, pero usando en este caso múltiples "Eigen-espacios", es decir, uno por cada expresión (clase) en lugar de un único espacio como se suele realizar en el reconocimiento facial. A este método se le denominó Eigenexpressions para indicar que sus fundamentos son similares al método de *Eigenfaces*, pero con diferencias relevantes; en particular, la construcción de múltiples subespacios durante la fase de entrenamiento, denominados "face spaces", y la integración de un clasificador para procesar los errores obtenidos en la reconstrucción de la expresión en cada uno de los subespacios.

5.2. Aproximación propuesta: Eigenexpressions

Una manera sencilla de aplicar el método *Eigenfaces* al problema del reconocimiento de expresiones, consiste en reemplazar las etiquetas de identificación del sujeto en el conjunto de entrenamiento X por otras que representen la expresión facial del sujeto.

Desafortunadamente, este método adolece de un problema de superposición de información. Hay que considerar el hecho de que las distancias en el "face space" no son exclusivas a la expresión, sino que tanto la similitud de la cara como la de la expresión facial pueden contribuir en la función del cálculo de la distancia. Una solución a este problema consiste en la substracción de la expresión neutral a todas las muestras de los conjuntos de entrenamiento y de test con el objetivo de eliminar dependencias entre sujetos. Sin embargo, esta estrategia sólo es posible si el sujeto ha sido previamente identificado y se dispone de su expresión neutral.

Una extensión más natural y general al método original de *Eigenfaces* para el reconocimiento de expresiones faciales consiste en el uso de múltiples "Eigenspaces". En primer lugar, el conjunto de entrenamiento X se divide en s conjuntos disjuntos $X_1, X_2, \ldots, X_s$ (uno por clase). A continuación PCA se aplica de forma separada a cada subconjunto X_i con el fin de computar s subespacios diferentes con sus correspondientes matrices de proyección $W_1, W_2, \ldots, W_s$. A continuación, la clasificación de una nueva muestra se realiza mediante el cálculo de la distancia euclídea de la muestra a cada uno de los subespacios, escogiendo el subespacio cuya distancia es la menor y asignando la etiqueta correspondiente a la nueva muestra. Esta aproximación ha sido aplicada con éxito en problemas de reconocimiento facial en [4], con mejoras de un 20 % en la precisión sobre el método tradicional de *Eigenfaces*.

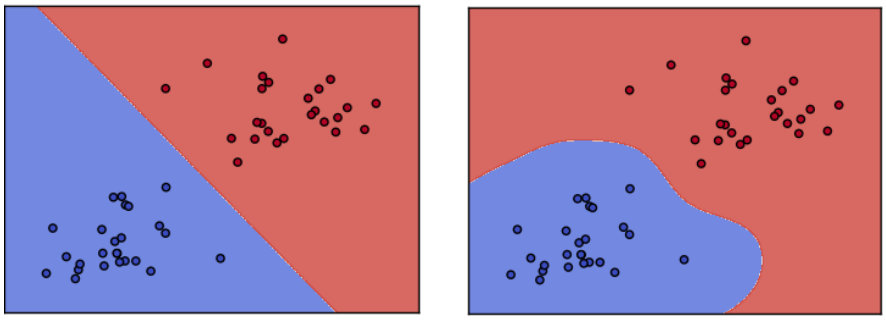
El criterio de la distancia mínima expuesto anteriormente asume una separación lineal entre las clases, con fronteras lineales definidas por distancias iguales a los subespacios. Este enfoque es mostrado en la figura 5.1(a), usando en este caso por simplicidad un espacio bidimensional.

No obstante, es posible tomar decisiones más robustas si las distancias a todos los subespacios son evaluadas de forma simultánea por un clasificador para determinar una frontera no lineal. Este enfoque es mostrado en la figura 5.1(b), mediante la definición de una frontera más compleja que la lineal. El método propuesto en Eigenexpressions toma esta aproximación en consideración.

En los dos siguientes subapartados se describirá el diseño de las fases de entrenamiento y reconocimiento implementadas en Eigenexpressions.

5.2.1. Diseño de la fase de entrenamiento

En Eigenexpressions el entrenamiento se realiza en dos etapas. La primera de ellas reproduce la extensión natural sobre el método estándar de Eigenfaces para el reconocimiento de expresiones faciales descrita en la aproximación propuesta. En este caso, el conjunto de entrenamiento $X = \{(x_i, l_i) | i = 1, \ldots, p\}$ se utilizó para el cálculo de los subespacios y sus respectivas matrices de proyección. Se consideraron seis subespacios ($s = 6$), uno para cada una de las seis emociones básicas: Alegría,



(a) *Separación lineal entre clases* **(b)** *Separación no lineal entre clases*

Figura 5.1: Fronteras de separación entre clases

Miedo, Tristeza, Asco, Ira y Sorpresa. Para calcular las proyecciones en cada uno de los subespacios $(W_1, W_2 \dots W_s)$, cada clase fue considerada de forma separada. Para cada clase c_j se calculó la matriz de proyección W_j usando exclusivamente PCA sobre las muestras del conjunto X de entrenamiento que pertenecen a la case c_j. Este proceso se exhibe en la figura 5.2.

En la imagen 5.3 se muestran, a título ilustrativo, seis ejemplos de unas *Eigenfaces* calculadas y reducidas mediante PCA para cada una de las emociones básicas, en escala de grises y con una resolución de 250 x 250 píxeles.

En la segunda etapa de la fase de entrenamiento se implementó y se entrenó un clasificador con las mismas muestras usadas en la primera etapa. Para cada muestra $x_i \in X$, se construyó un vector de características d_i con la distancia euclídea de x_i a cada uno de los s subespacios, representado como $d_i = \{d_{i,j} | j = 1, 2, \dots, s\}$, siendo $d_{i,j}$ el error cuadrático medio de la reconstrucción de la imagen x_i en el subespacio j. Esta segunda etapa de entrenamiento se muestra en la figura 5.4.

Asumiendo que la media de todas las imágenes es substraída de todos los vectores x_i durante el proceso de normalización, este error puede ser formalmente definido como $d_{i,j} = ||x_i - x_i W_j^T W_j||$.

Por último, las etiquetas de cada imagen se usaron para la construcción del conjunto de resultados $\{(d_i, l_i) | i = 1, \dots, p\}$, el cual sirvió de entrada en la etapa de entrenamiento del clasificador implementado en Eigenexpressions.

5.2.2. Diseño de la fase de reconocimiento

En la figura 5.5 se ilustra el proceso de reconocimiento de expresiones. En esta etapa, dada una nueva imagen de entrada se calculan los s vectores de distancia d_i de la imagen a cada uno de los subespacios. Estos vectores constituyen la entrada del clasificador construido en la segunda etapa de entrenamiento para predecir la expresión apropiada.

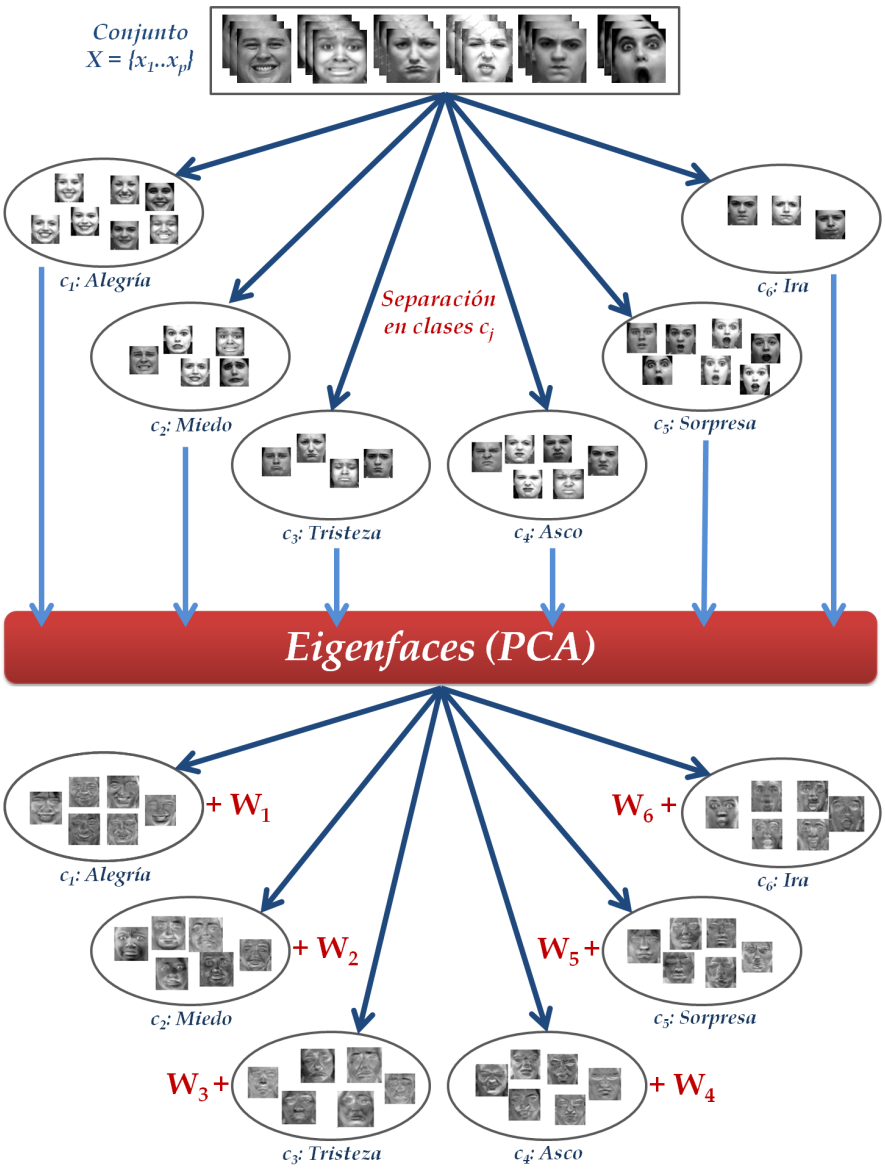
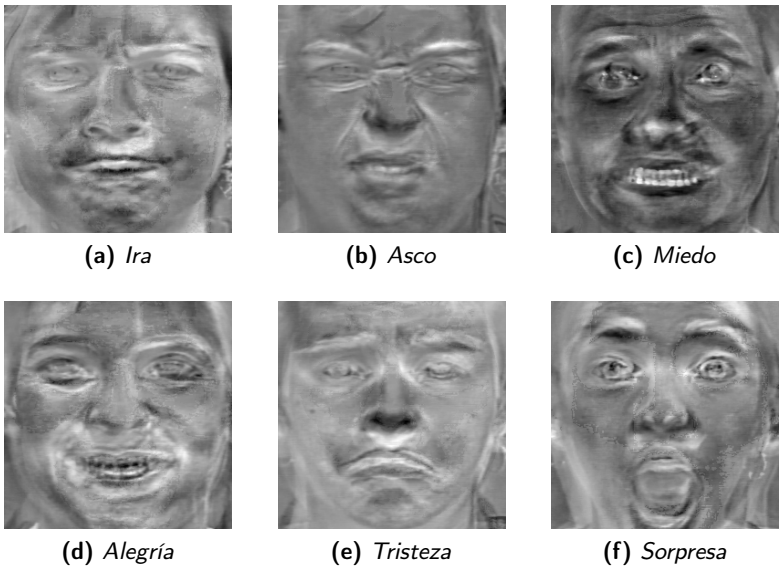


Figura 5.2: Primera etapa de entrenamiento en Eigenexpressions

5.3. Configuración de Eigenexpressions

En los dos subsiguientes subapartados se detallará la configuración de Eigenexpressions para poder llevar a cabo la experimentación y evaluación del método para la detección de expresiones faciales. En primer lugar, se describirá la base de datos empleada para la construcción de los clasificadores. A continuación,



(a) *Ira* **(b)** *Asco* **(c)** *Miedo*

(d) *Alegría* **(e)** *Tristeza* **(f)** *Sorpresa*

Figura 5.3: Ejemplo de *Eigenfaces* obtenidas y reducidas mediante PCA para cada una de las seis emociones básicas

se finalizará con el detalle de la configuración experimental empleada para la posterior evaluación el método.

5.3.1. Base de datos empleada

Eigenexpressions fue construido y evaluado utilizando la base de datos Cohn-Kanade+ (*The Extended Cohn-Kanade Dataset*) [115]. Esta base de datos incluye 593 secuencias de imágenes de 123 sujetos. Cada secuencia contiene entre 10 y 60 imágenes en una sucesión de fotogramas, comenzando la secuencia desde su expresión neutral hasta el *clímax* de la expresión correspondiente a una de las seis emociones básicas. Para una gran parte de las secuencias de imágenes, aunque no para la totalidad de las incluidas en la base de datos, se incluye también una etiqueta validada de la emoción representada por el sujeto. Únicamente la última imagen de cada secuencia validada, correspondiente al *clímax* de la expresión, se consideró para la construcción y evaluación de Eigenexpressions, lo que dio lugar a un conjunto de 123 sujetos con entre 1 a 6 imágenes por individuo, cada una representando una expresión diferente.

Sobre esta selección de imágenes se aplicó una implementación estándar del algoritmo Viola-Jones [184] para detectar y recortar la cara del sujeto con el objetivo de que todas las imágenes fueran lo más homogéneas posibles. Por último, para hacer que el sistema fuera invariante a la intensidad de la iluminación del sujeto, los histogramas de las imágenes recortadas fueron ecualizados.

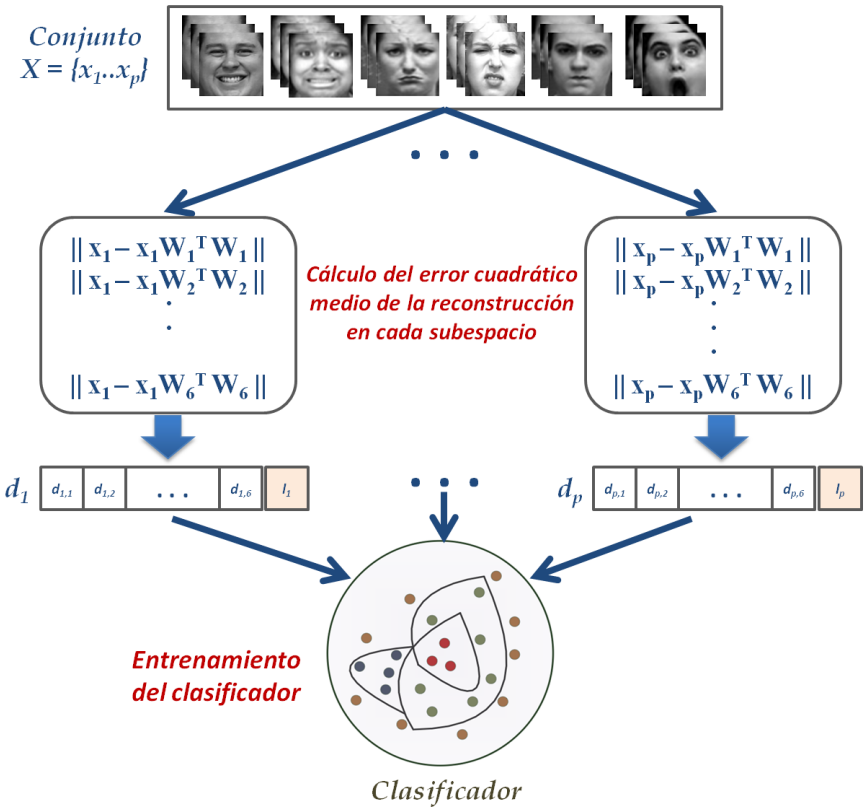


Figura 5.4: Segunda etapa de entrenamiento en Eigenexpressions

5.3.2. Configuración experimental

Para evaluar el rendimiento de Eigenexpressions y compararlo con el reconocimiento de expresiones mediante el método estándar de *Eigenfaces* se crearon dos implementaciones diferentes, realizándose sobre cada una de ellas una experimentación típica de tipo entrenamiento / test. En primer lugar, ambos sistemas se entrenaron con el mismo conjunto de muestras de entrenamiento. Seguidamente se ejecutaron ambos clasificadores sobre el mismo conjunto de test, realizando un estudio comparativo de sus resultados. En cada ejecución del experimento el número de muestras de entrenamiento por expresión (w) fueron inicialmente fijadas, analizándose diferentes valores para el parámetro w ($w = 9, \ldots, 17$). Con el fin de mantener la consistencia entre las medidas de los errores de reconstrucción entre diferentes subespacios, se empleó el mismo número de muestras de entrenamiento para cada una de las seis expresiones faciales. Por último, para aumentar la fiabilidad de los resultados, todos los experimentos se ejecutaron cuatro veces y sus resultados promediados.

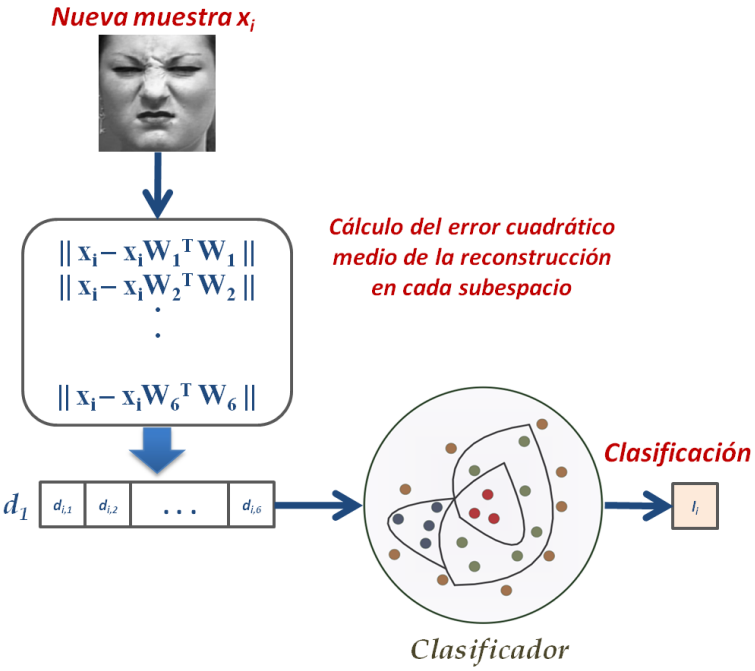


Figura 5.5: Etapa de reconocimiento en Eigenexpressions

5.4. Evaluación de Eigenexpressions

Para la determinación del método de clasificación utilizado en la segunda etapa de entrenamiento de Eigenexpressions se evaluaron diferentes estrategias de clasificación. Aunque el uso de diferentes esquemas de clasificación no representaron un impacto significativo en los experimentos, el modelo LMT [105] mostró un mayor rendimiento en los resultados frente a otras estrategias, motivo por el que se decidió usar este modelo sobre la base de datos Cohn-Kanade+. En las tablas 5.1 y 5.2 se muestra un resumen de los tres esquemas de clasificación que demostraron un comportamiento mejor en el proceso de reconocimiento de expresiones faciales, en términos de exactitud media alcanzada y área media bajo la curva ROC (característica operativa del receptor o *Receiver Operating Characteristic*). En las tablas se aprecia que el esquema LMT mostró un rendimiento superior al perceptrón multicapa y a la SVM para cualquier tamaño de w. Los resultados para otros valores intermedios de w no se tomaron en consideración por ser consistentes con los mostrados en la tabla.

En la tabla 5.3 se muestra una comparativa de la exactitud global de Eigenexpressions frente al método estándar de *Eigenfaces* para varios tamaños representativos de conjuntos de entrenamiento, con tamaños para $w = 9$, 14 y 17. De modo equivalente al análisis de los esquemas de clasificación evaluados, los

w	**LMT**	**Perceptrón multicapa**	**SVM**
9	76,3 %	68,6 %	68,4 %
14	79,1 %	73,2 %	69,9 %
17	79,6 %	72,5 %	72,7 %

Tabla 5.1: Comparativa de la exactitud global (*accuracy*) entre diversos esquemas de clasificación en Eigenexpressions para diferentes tamaños de w

w	**LMT**	**Perceptrón multicapa**	**SVM**
9	0,9393	0,9293	0,8920
14	0,9440	0,9343	0,9037
17	0,9607	0,9557	0,9263

Tabla 5.2: Comparativa del área ROC entre diversos esquemas de clasificación en Eigenexpressions para diferentes tamaños de w

resultados para otros valores intermedios de w no se tomaron en consideración por ser consistentes con la información mostrada en la tabla. A la vista de los resultados, se puede concluir que, en promedio, el método de *Eigenfaces* falla en la clasificación una por cada tres muestras, mientras que Eigenexpressions falla una por cada cinco, obteniéndose una ganancia media de un 16,9 % para $w = 14$ con respecto a *Eigenfaces*.

w	**Eigenfaces**	**Eigenexpressions**	**Ganancia**
9	66,3 %	76,3 %	15,1 %
14	69,4 %	81,1 %	16,9 %
17	69,9 %	79,5 %	13,7 %

Tabla 5.3: Comparativa de la exactitud global (*accuracy*) de *Eigenfaces* frente a Eigenexpressions en el reconocimiento de expresiones faciales para diferentes tamaños de w

Los resultados en términos de los valores de cobertura por clase entre ambos métodos se analizan en las figuras 5.6, 5.7 y 5.8 para distintos tamaños de w. En ellas se observa que Eigenexpressions supera de forma consistente y considerable al método estándar de reconocimiento mediante *Eigenfaces* para las emociones Sorpresa, Alegría y Asco. Respecto a la emoción Miedo, *Eigenfaces* obtiene mejores resultados, excepto para $w = 14$ donde existe un empate entre ambos enfoques. En cuanto al rendimiento de las dos clases restantes, Ira y Tristeza, éste dependió de los distintos valores de w, no mostrando ninguno de los dos métodos presentados una ganancia significativamente mayor frente al otro. La consistencia en cuanto a rendimiento de ambos métodos en las clases correspondientes a Sorpresa, Alegría, Asco y Miedo, sugiere que ambos métodos podrían ser combinados en aproximaciones híbridas.

Por último, merece la pena destacar otro resultado obtenido y relacionado con la contribución de la etapa final de clasificación en Eigenexpressions, la cual es

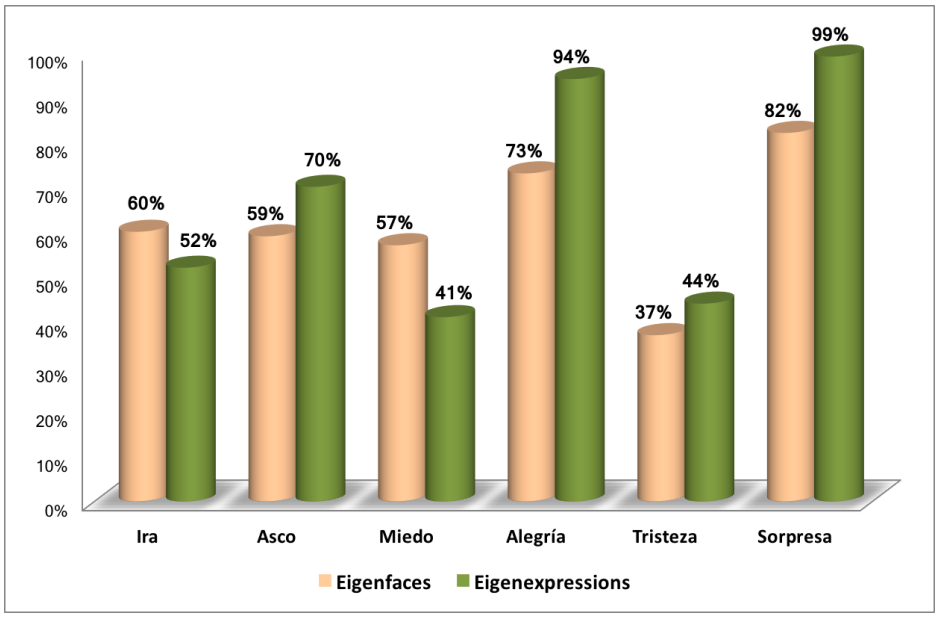


Figura 5.6: Comparativa de los valores de cobertura por clase entre *Eigenfaces* y Eigenexpressions para un tamaño de entrenamiento $w = 9$

usada con el objetivo de incrementar la robustez del método al permitir disponer de fronteras no lineales entre las clases. Para evaluar esta contribución, se ejecutó el método descrito en el apartado 5.2, para el reconocimiento de expresiones faciales mediante *Eigenfaces* con múltiples subespacios, sobre el mismo conjunto de entrenamiento y test. La exactitud global (*accuracy*) de esta aproximación comparada con Eigenexpressions se muestra en la tabla 5.4. En ella se observan pequeñas ganancias en el rango 1–4 % en favor de Eigenexpressions para diferentes tamaños en los conjuntos de entrenamiento.

w	**Múltiples subespacios**	**Eigenexpressions**	**Ganancia**
9	73,4 %	76,3 %	4,0 %
14	79,9 %	81,1 %	1,5 %
17	78,7 %	79,5 %	1,0 %

Tabla 5.4: Comparativa de la exactitud global (*accuracy*) del método de múltiples subespacios frente a Eigenexpresssions para diferentes tamaños de entrenamiento w

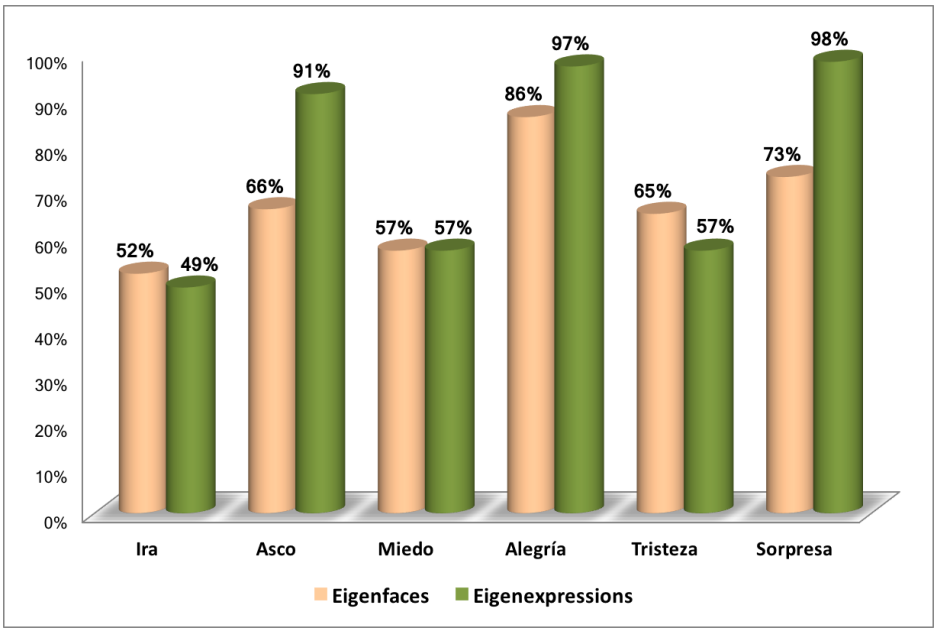


Figura 5.7: Comparativa de los valores de cobertura por clase entre *Eigenfaces* y Eigenexpressions para un tamaño de entrenamiento $w = 14$

5.5. Extensión de Eigenexpressions mediante máscaras de expresiones faciales

Fundamentado en la idea de que no todas las áreas, puntos o píxeles que definen una cara contienen información relevante para la detección de una determinada expresión facial y que en cada expresión ciertas zonas o píxeles relevantes pueden diferir, es factible un planteamiento a través del cual sea posible crear una máscara que elimine o asigne menor importancia a aquellos píxeles que presentan una alta variabilidad y que no contribuyen en la formación de la expresión facial. Desde esta perspectiva, las máscaras de expresiones faciales excluirán o darán menor importancia a las partes de la cara que no contribuyen a la expresión facial, mientras que, por el contrario, a las zonas más relevantes les asignará un mayor peso en el proceso de la clasificación.

En los siguientes subapartados se describirá la creación de las máscaras para el conjunto de las seis expresiones faciales consideradas, el entrenamiento del clasificador mediante esta nueva aproximación y el proceso final de reconocimiento de expresiones faciales.

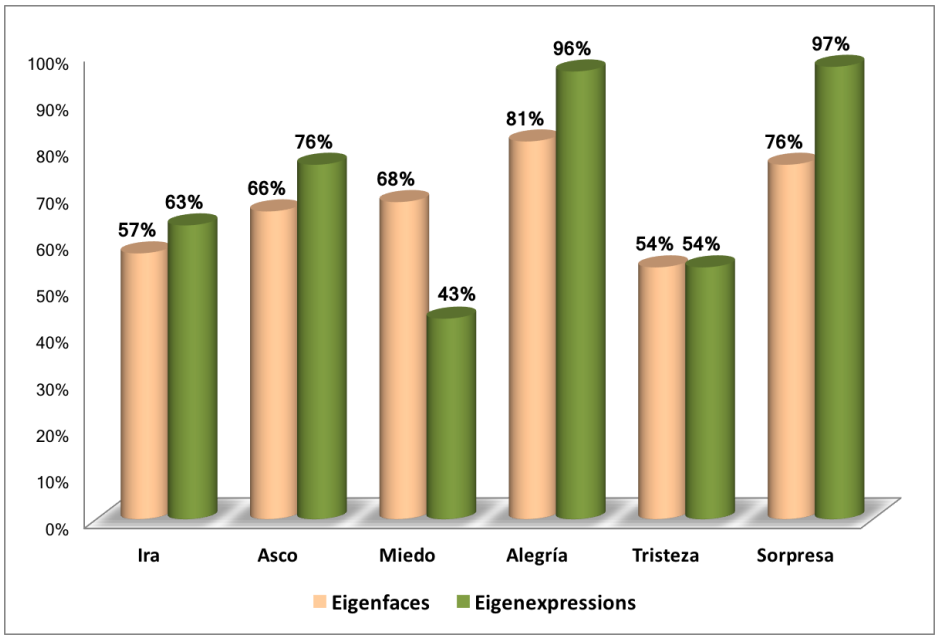


Figura 5.8: Comparativa de los valores de cobertura por clase entre *Eigenfaces* y Eigenexpressions para un tamaño de entrenamiento $w = 17$

5.5.1. Formulación y creación de las máscaras

De modo similar al estudio realizado en [71] en el que se utiliza una aproximación de máscaras faciales basadas en el método de *Eigenfaces* para la detección de dos expresiones faciales (neutral y enfado), utilizando el método Eigenexpressions descrito en los apartados anteriores, se ha creado una máscara para cada una de las seis posibles expresiones faciales con el propósito de caracterizar las zonas más relevantes de la expresión. Cada máscara es generada a partir del "face space" correspondiente a una expresión c_j determinada, proyectando sobre el subespacio de la expresión, por un lado, el conjunto $Q = \{x_1, \ldots, x_q\}$ de todas aquellas imágenes de X que corresponden a la expresión facial del "face space" y, por otro, el conjunto $R = \{x_1, \ldots, x_r\}$ del resto de imágenes $x_m \in X$ que no corresponden a la expresión del subespacio. El objetivo de las proyecciones es poder computar los errores medios de reconstrucción de las imágenes de la clase c_j, así como del resto de las imágenes que no pertenecen a dicha clase. El cálculo de los errores cuadráticos medios se describe en las ecuaciones 5.1 y 5.2.

$$d_{Qj} = \frac{1}{q} \sum_{k=1}^{q} |x_k - x_k \mathrm{W}_j^T \mathrm{W}_j| \ ; \ x_k \in Q \tag{5.1}$$

$$d_{Rj} = \frac{1}{r} \sum_{m=1}^{r} |x_m - x_m \mathrm{W}_j^T \mathrm{W}_j| \ ; \ x_m \in R \tag{5.2}$$

donde d_{Qj} representa el error cuadrático medio de reconstrucción sobre el "face space" W_j de todas las imágenes x_k del conjunto de entrenamiento Q (subconjunto de imágenes de X que pertenecen a la clase c_j), mientras que d_{Rj} constituye el error cuadrático medio de reconstrucción de R (resto de las imágenes que no pertenecen a la clase c_j). La suma de las muestras q y r constituyen el total de las imágenes del conjunto de entrenamiento X.

Para el cálculo de la máscara M_j para cada clase c_j, el conjunto de muestras de entrenamiento se divide en dos grupos disjuntos: 1) las imágenes x_k que pertenecen a la clase c_j (grupo Q); 2) el resto de imágenes x_m que no pertenecen a la clase c_j (grupo R). Sobre cada uno de estos conjuntos se obtienen los errores cuadráticos medios de reconstrucción d_{Rj} y d_{Qj} como resultado de su proyección y reconstrucción sobre el "face space" j, obteniéndose la máscara para la clase c_j como la diferencia entre el error medio de reconstrucción de las imágenes que corresponden a la clase y el error medio de las que no corresponden a la misma, dividida por el error cuadrático medio d_j de la proyección del conjunto total de imágenes de X sobre el subespacio j. En las ecuaciones 5.3 y 5.4 se expresan formalmente los cálculos anteriormente descritos. Por último, los valores negativos de la máscara M_j son establecidos a cero y su resultado normalizado.

$$d_j = \frac{1}{n}\sum_{i=1}^{n} |x_i - x_i \mathrm{W}_j^T \mathrm{W}_j| \ ; \ x_i \in X \tag{5.3}$$

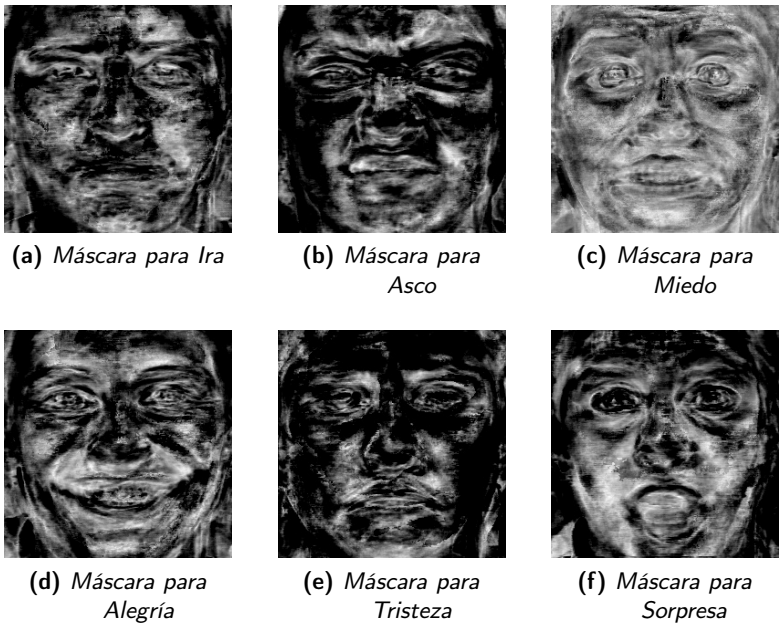
$$M_j = \frac{d_{Rj} - d_{Qj}}{d_j} \tag{5.4}$$

En la figura 5.9 se muestran seis ejemplos de máscaras para cada una de las expresiones faciales consideradas. Los píxeles negros de las máscaras representan las zonas de la cara con alta variabilidad y que no contribuyen a la expresión facial, es decir son partes del rostro innecesarias para la clasificación de la expresión. Por el contrario, las zonas más claras indican rasgos comunes entre sujetos para una misma expresión facial y, por lo tanto, relevantes para su reconocimiento.

5.5.2. Entrenamiento mediante máscaras

En este caso, de modo similar al método descrito en la aproximación propuesta para Eigenexpressions, el entrenamiento se realiza en tres etapas: 1) separación del conjunto de las muestras de entrenamiento en diferentes clases para cada expresión facial considerada y creación de los "face spaces"; 2) construcción de las máscaras M_j, una por cada subespacio y expresión facial; 3) proyección de las imágenes sobre todos los subespacios, aplicación de las máscaras y cálculo de los vectores de características d_i con la distancia euclídea a cada uno de los subespacios tras la aplicación de la máscara y construcción del clasificador.

Es decir, al proceso de entrenamiento propuesto en Eigenexpressions se le ha añadido una etapa adicional consistente en la creación de una máscara M_j para cada expresión facial considerada. Estas máscaras son aplicadas en la última etapa de la fase de entrenamiento para el cálculo del vector de distancias a cada subespacio.



(a) *Máscara para Ira* **(b)** *Máscara para Asco* **(c)** *Máscara para Miedo*

(d) *Máscara para Alegría* **(e)** *Máscara para Tristeza* **(f)** *Máscara para Sorpresa*

Figura 5.9: Ejemplo de las máscaras obtenidas con *Eigenfaces* para cada una de las seis emociones básicas

5.5.3. Reconocimiento mediante máscaras

Dada una nueva cara x_i con una expresión facial determinada, el proceso de reconocimiento se realizará proyectando la imagen sobre cada uno de los "face spaces" W_j creados durante la fase de entrenamiento. Seguidamente se le aplicará la máscara correspondiente a cada subespacio con el objetivo de potenciar las zonas comunes a la expresión y minimizando o eliminando aquellas que presentaron una alta variabilidad en la creación de la máscara y que pudieran no ser relevantes en la detección de la expresión facial, actuando la máscara como si de un filtro se tratara. Tras este "filtrado" se obtendrá el vector de distancias d_j a cada subespacio como una medida normalizada del error de reconstrucción, información que será pasada al clasificador para la predicción de la expresión facial apropiada.

5.6. Evaluación de la extensión mediante máscaras

En la tabla 5.5 se muestra una comparativa de la exactitud global (*accuracy*) de la aproximación basada en máscaras frente al método de Eigenexpressions descrito en los apartados anteriores, para diferentes tamaños representativos para el conjunto de entrenamiento ($w = 9, 14$ y 17). De modo similar al análisis

comparativo realizado entre Eigenexpressions y *Eigenfaces*, los resultados para valores intermedios de w no se reflejan en las tablas por ser consistentes con los obtenidos para los valores de w mostrados. En esta tabla se observa que la aproximación basada en máscaras presenta un rendimiento medio inferior cercano al 4 % con respecto a los resultados obtenidos con Eigenexpressions.

w	**Eigenexpressions**	**Máscaras**	**Ganancia**
9	76,3 %	73,5 %	-3,7 %
14	81,1 %	76,3 %	-5,9 %
17	79,5 %	76,7 %	-3,5 %

Tabla 5.5: Comparativa de la exactitud global (*accuracy*) de la aproximación basada en máscaras frente al método Eigenexpressions para diferentes tamaños de entrenamiento w

Por el contrario, en la tabla 5.6 se aprecia una ganancia significativa de la aproximación basada en máscaras para el reconocimiento de expresiones faciales frente al método estándar de *Eigenfaces*, con valores medios de mejora cercanos al 10 % para diferentes tamaños de entrenamiento, aunque sigue presentando un rendimiento inferior a Eigenexpressions, donde las mejoras de esta última aproximación frente al método estándar de Eigenfaces se sitúan entre el 14 % y el 17 %, según los datos mostrados previamente en tabla 5.3.

w	**Eigenfaces**	**Máscaras**	**Ganancia**
9	66,3 %	73,5 %	10,9 %
14	69,4 %	76,3 %	9,9 %
17	69,9 %	76,7 %	9,7 %

Tabla 5.6: Comparativa de la exactitud global (*accuracy*) de la aproximación basada en máscaras frente al método estándar de *Eigenfaces* para diferentes tamaños de entrenamiento w

Por último, como se muestra en la tabla 5.7, tampoco se obtuvieron mejoras con la aproximación basada en máscaras frente al método estándar de *Eigenfaces* mediante el uso de múltiples subespacios.

w	**Múltiples subespacios**	**Máscaras**	**Ganancia**
9	73,4 %	73,5 %	0,1 %
14	79,9 %	76,3 %	-4,5 %
17	78,7 %	76,7 %	-2,5 %

Tabla 5.7: Comparativa de la exactitud global (*accuracy*) de la aproximación basada en máscaras frente al método estándar de *Eigenfaces* mediante múltiples subespacios para diferentes tamaños de entrenamiento w

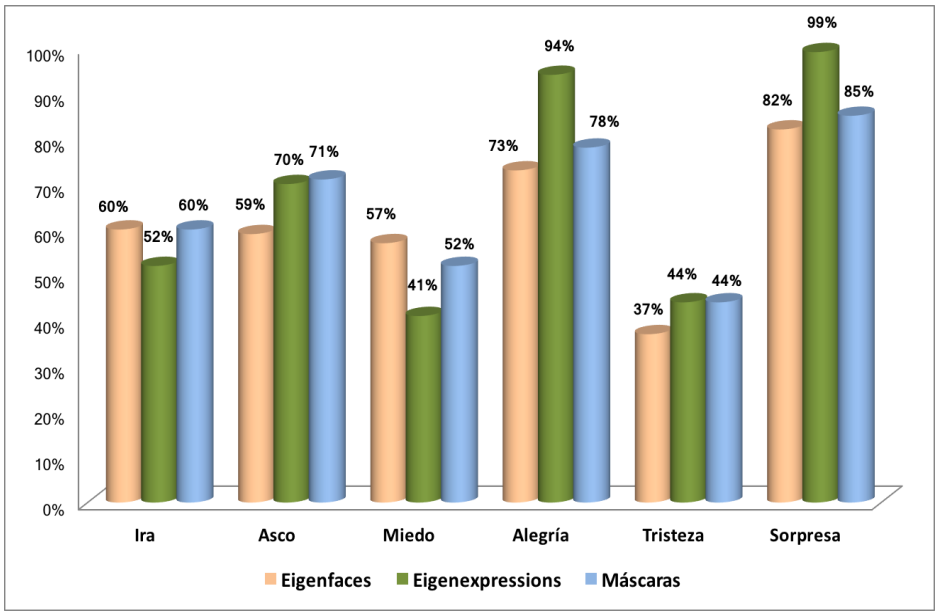


Figura 5.10: Comparativa de los valores de cobertura por clase entre *Eigenfaces*, Eigenexpressions y la aproximación basada en máscaras, para un tamaño de entrenamiento con $w = 9$

Los resultados en términos de valores de cobertura por clases entre los tres métodos evaluados, *Eigenfaces*, Eigenexpressions y máscaras, se comparan en las figuras 5.10, 5.11 y 5.12 para diferentes tamaños de w.

De modo similar al método propuesto por Eigenexpressions, donde éste superaba de forma consistente y considerable al método estándar de *Eigenfaces* en el reconocimiento de las emociones correspondientes a la Sorpresa, la Alegría y el Asco, la aproximación mediante máscaras superó, en general, al método estándar de *Eigenfaces* en el reconocimiento de la Sorpresa, el Asco y la Ira. Del conjunto de estas cuatro expresiones detectadas con mayor precisión por ambos métodos, Eigenexpressions presentó una tasa de aciertos superior con respecto a la aproximación mediante máscaras en las expresiones de Sorpresa, con un 98 % de media en Eigenexpressions frente al 86 % con máscaras; y al Asco con un 79 % en Eigenexpressions de media frente al 71 % con máscaras. Sin embargo, la emoción Ira fue detectada, en todos los casos, con mayor precisión y de forma más consistente con el método basado en máscaras, con una media de un 62 % frente a un 55–56 % en Eigenexpressions y Eigenfaces.

Sobre el resto de emociones, los resultados no aportaron evidencias que demostraran que un método funcionara mejor con respecto a los otros dos. El Miedo presentó unos resultados más estables en la aproximación basada en *Eigenfaces* entre diferentes tamaños de entrenamiento w. Sin embargo la Tristeza mostró grandes variaciones entre métodos y tamaños de entrenamiento, no pudiéndose

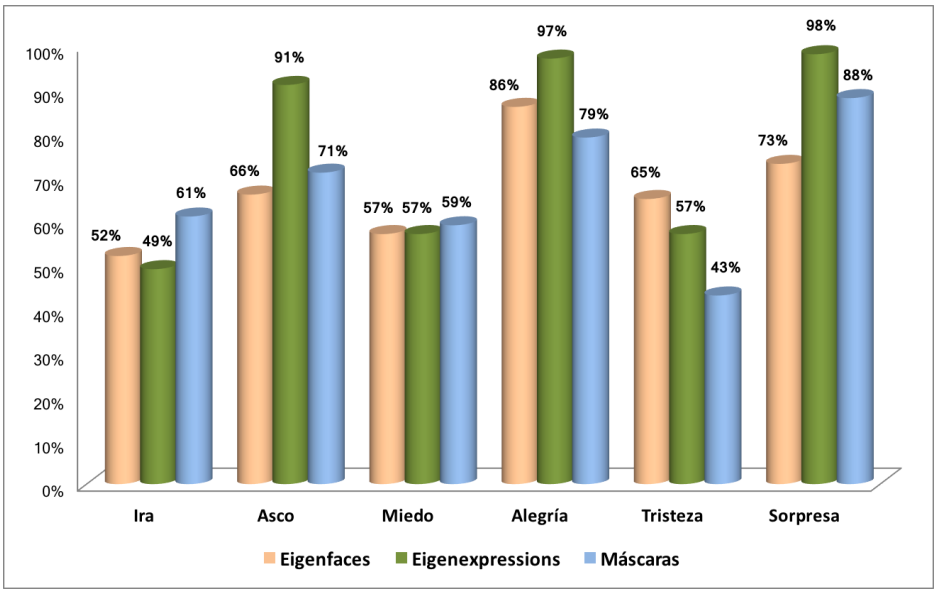


Figura 5.11: Comparativa de los valores de cobertura por clase entre *Eigenfaces*, Eigenexpressions y la aproximación basada en máscaras, para un tamaño de entrenamiento con $w = 14$

concluir que uno de los métodos fuera más adecuado que el resto en la detección de esta emoción.

Estos resultados sugieren, de nuevo, que una aproximación híbrida en el que se empleen diversos métodos podría ofrecer tasas de acierto más elevadas que el uso de aproximaciones individuales.

5.7. Conclusiones

El método propuesto y desarrollado en Eigenexpressions como aproximación holística basada en la apariencia para el reconocimiento de expresiones faciales, así como la extensión propuesta para la detección mediante máscaras faciales por expresión, fueron evaluados sobre el elenco de imágenes contenidas en la base de datos Cohn-Kanade+ y, en consecuencia, el reconocimiento de la expresión facial fue realizado sobre el conjunto de las seis emociones básicas contenidas en la misma. No obstante, a pesar de que Eigenexpressions fue entrenado y evaluado únicamente con estas seis emociones primarias, el sistema es fácilmente extensible a cualquier conjunto de expresiones simplemente incluyendo en el conjunto de etiquetas los nuevos descriptores para cada nueva emoción, aunque el rendimiento del sistema tras la extensión de nuevas expresiones debería ser analizado en cada contexto específico, debido a que diferentes condiciones podrían presentar un potencial impacto en su rendimiento, como podría ser la distancia entre los diferentes subespacios, al fundamentarse, tanto Eigenexpressions como

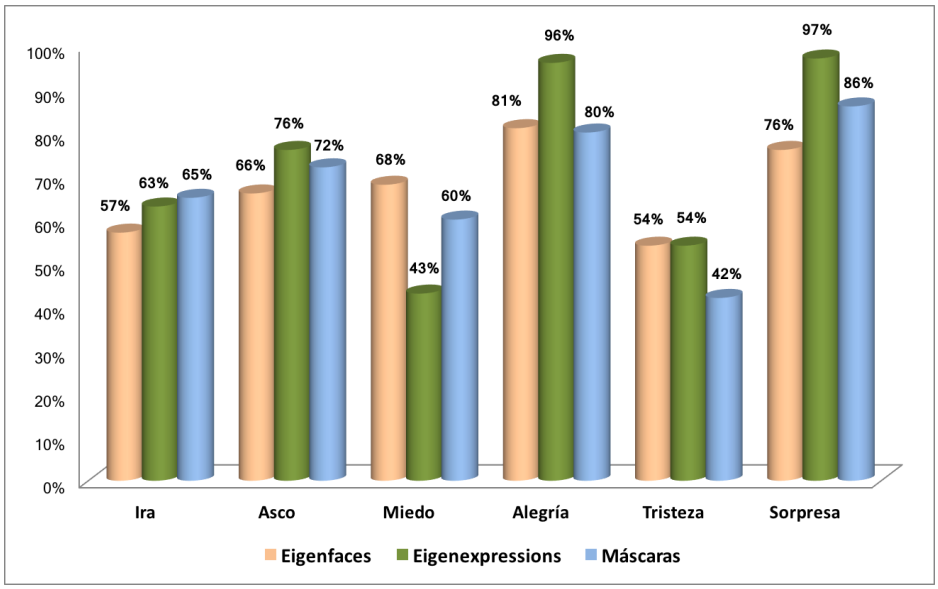


Figura 5.12: Comparativa de los valores de cobertura por clase entre *Eigenfaces*, Eigenexpressions y la aproximación basada en máscaras, para un tamaño de entrenamiento con $w = 17$

su extensión mediante máscaras, en la construcción de un subespacio (*eigenspace*) por expresión y en el uso de un clasificador cuya salida se basa en la distancia de la muestra a cada uno de estos subespacios.

Con Eigenexpressions, así como con su extensión basada en máscaras faciales por expresión, se logró una mejora substancial con respecto al método estándar *Eigenfaces* para el reconocimiento de expresiones faciales, usando para ello el cálculo de la distancia euclídea a cada subespacio. No obstante, el uso de funciones alternativas para el cálculo de las distancias y su evaluación son objetivos a abordar en el futuro.

Por otro lado, un mayor número de muestras disponibles en la fase de entrenamiento podría aportar, previsiblemente, potenciales beneficios adicionales. Este escenario permitiría dividir el conjunto de entrenamiento $X = \{(x_i, l_i)|i = 1, \ldots, p\}$ en dos conjuntos disjuntos $X_1 = \{(x_i, l_i)|i = 1, \ldots, g\}$ y $X_2 = \{(x_i, l_i)|i = g + 1, \ldots, p\}$, usando cada uno de ellos en cada una de las dos etapas del proceso de entrenamiento, minimizando, de esta forma, el problema de superposición de información, ya que, de modo similar a como sucede con el método de *Eigenfaces*, es necesario considerar el hecho de que las distancias en los "face spaces" no son exclusivas a la expresión, sino que tanto la similitud de la cara como la de la expresión facial pueden contribuir en la función del cálculo de la distancia. Una solución a este problema consistiría en la substracción de la expresión neutral a todas las muestras de los conjuntos de entrenamiento y de test con el objetivo de eliminar dependencias entre sujetos. Sin embargo, esta estrategia sólo es posible si

el sujeto ha sido previamente identificado y se dispone de su expresión neutral, lo que no siempre es factible.

Aunque la base de datos empleada en la evaluación de Eigenexpressions, así como su extensión mediante máscaras, contiene imágenes tomadas bajo similares condiciones lumínicas, es importante considerar que los métodos de reducción basados en PCA, como el utilizado en Eigenexpressions y en su extensión mediante máscaras, son altamente sensibles a cambios en la iluminación de la escena, circunstancia que podría reducir drásticamente el rendimiento global del sistema de reconocimiento. Aunque, en este sentido, se ha intentado reducir el impacto de la variación lumínica mediante la ecualización del histograma, otras posibles soluciones alternativas podrían explorarse en el futuro. En cualquier caso, es previsible que el uso de otros conjuntos de entrenamiento tomados bajo diferentes condiciones lumínicas tenga un efecto similar al método de construcción de subespacios lineales tridimensionales, como se describe en [22]; esto podría reducir la sensibilidad del método, como por ejemplo frente a cambios en la dirección de la luz.

A pesar de los resultados positivos obtenidos con las aproximaciones propuestas en este capítulo, el reconocimiento de expresiones faciales en humanos sigue siendo una tara compleja y no exenta de imprecisiones. Por ello, es factible considerar que su precisión podría ser mejorada si para el análisis se hubiera podido disponer de varias fuentes de información simultáneas, como por ejemplo características geométricas faciales, seguimiento de ojos, análisis de la voz, de la postura o de señales fisiológicas. En esta dirección, algunos sistemas de detección afectiva multimodal analizan simultáneamente tres o más fuentes de información [32].

En resumen, la simplicidad del método expuesto en Eigenexpressions y, por analogía, en su extensión mediante máscaras de expresión facial, hace que el sistema sea adecuado tanto para su aplicación de forma aislada como clasificador de expresiones, como en aproximaciones multimodales donde podrían aportar información sobre la expresión facial detectada y ser usados como fuentes de información complementarias junto a otros métodos basados en características diferentes [31, 32]. Adicionalmente, las substanciales diferencias existentes entre los resultados obtenidos por el método estándar de *Eigenfaces*, por Eigenexpressions y por su extensión mediante máscaras, sugieren que los tres métodos podrían ser considerados como fuentes de información *quasi* independientes, aspecto que abre la posibilidad a la combinación de estas aproximaciones para que, de forma simultánea, proporcionen un resultado que ayude en el reconocimiento de la expresión facial.

Capítulo 6

Detección emocional sobre señales fisiológicas: Análisis de MAHNOB-HCI

Resumen
En este último capítulo sobre aportaciones en el área de la detección emocional, se describirá el análisis de información de carácter fisiológico realizado a través de la selección y extracción de características sobre el conjunto de señales procedentes de electroencefalografía contenidas en la base de datos multimodal MAHNOB-HCI.

Contenidos

6.1. Análisis de la base de datos MAHNOB-HCI 86
6.2. Selección y extracción de características EEG 88
6.3. Clasificación y análisis de resultados 91
6.4. Conclusiones 97

Diversos estudios aseveran que la actividad fisiológica del sujeto puede tener un impacto significativo en su estado afectivo [159]. Mediciones y características extraídas de señales biológicas como la variabilidad del pulso cardíaco (HRV), la respuesta galvánica de la piel (GSR), la temperatura corporal, los patrones de respiración o las bandas de las señales EEG, han demostrado que existen correlaciones con el estado emocional [106, 148, 123, 125, 186, 95, 94].

En contraposición a los métodos y fuentes de información empleados en los capítulos anteriores para la detección de la expresión facial del usuario y predicción del correspondiente estado afectivo, con el propósito de explorar las posibilidades que pueden ofrecer otras fuentes de información alternativas para la detección del

estado afectivo subyacente, en este capítulo se describirá el análisis realizado sobre un conjunto de señales fisiológicas procedentes de electroencefalografía (EEG) con el mismo objetivo: la predicción del estado afectivo del usuario.

No obstante, es importante destacar que mientras las aproximaciones basadas en técnicas de visión artificial, como las empleadas los capítulos 4 y 5 son poco intrusivas para el usuario y razonablemente económicas, las técnicas basadas en la captura y análisis de señales fisiológicas suelen ser lo contrario: altamente intrusivas y dependientes de instrumental costoso y de precisión.

Como sucede en cualquier proyecto o sistema que desee evaluar el estado afectivo de un usuario, un requisito indispensable es disponer de un corpus para poder entrenar y evaluar el rendimiento de la implementación llevada a cabo. En el ámbito de las bases de datos públicas que incorporan señales fisiológicas anotadas con su correspondiente estado afectivo destacan MAHNOB-HCI [169], DEAP (*Database for Emotion Analysis using Physiological Signals*) [97], EMDB (*Emotional Movie Database*) [34] y, más recientemente, DECAF (*Multimodal Dataset for Decoding Affective Physiological Responses*) [1]. Se escogió MAHNOB-HCI por ser, en la fecha de realización de este estudio (2011), una de las que aportaba un mayor número de canales de información así como de sujetos participantes.

De forma previa a la descripción del trabajo realizado para la selección y extracción de características fisiológicas basadas en señales procedentes de EEG, su clasificación y posterior evaluación de los resultados, se procederá con la descripción de la base de datos empleada para la extracción y construcción del conjunto de datos que pudieran resultar relevantes para un sistema de predicción del estado afectivo de un usuario: MAHNOB-HCI.

6.1. Análisis de la base de datos MAHNOB-HCI

MAHNOB-HCI [169] es una base de datos multimodal que contiene los registros de la respuesta afectiva y espontánea de sujetos expuestos a diferentes estímulos de carácter emocional a través de fragmentos de películas y grabaciones de vídeo, convenientemente etiquetados por sus participantes tras la visualización del estímulo. La base de datos, disponible de forma gratuita para la comunidad académica interesada en áreas de investigación como puede ser la relativa a la computación afectiva, la psicológica o la publicitaria, entre otras, proporciona información de cinco fuentes plenamente sincronizadas: vídeos de la cara y del cuerpo tomados desde diferentes ángulos, señales de audio, datos sobre el seguimiento de los ojos y señales fisiológicas del sistema nervioso periférico y central.

El corpus reúne datos de 27 participantes, de los cuales 11 son hombres y 16 mujeres. Los estados afectivos recogidos son: Neutral, Ansiedad, Diversión, Tristeza, Alegría, Asco, Ira, Sorpresa y Miedo. En la tabla 6.1 se detallan las emociones reportadas por los sujetos participantes a través de la visualización de 20 fragmentos de vídeo reproducidos de modo aleatorio y el número de muestras disponibles en cada una de estas categorías.

Las señales grabadas y entregadas en MAHNOB-HCI son:

1. Grabaciones de las expresiones faciales y las posturas corporales adoptadas, tomadas con 6 cámaras a una velocidad de 60 fotogramas por segundo y a una resolución de 780 x 580 píxeles. Únicamente la cara frontal fue grabada con una cámara en color. Las cinco restantes fueron recogidas en blanco y negro. Las tomas fueron realizadas desde distintos ángulos: frontal, parte superior de la pantalla, lateral inferior izquierdo, lateral inferior derecho, perfil y con un objetivo angular desde la parte superior de la pantalla para la captura de la cara, el torso y los brazos.

2. Información del seguimiento ocular de los usuarios mientras visionaban los vídeos. En concreto 24 características sobre las coordenadas de los ojos, tomadas a una velocidad de 60 Hz proyectadas sobre una matriz de resolución 1280 x 800 píxeles; 6 características sobre el diámetro de la pupila; 4 sobre la distancia de los ojos al sensor y 4 sobre el parpadeo.

3. Grabaciones de audio, compuestas principalmente de interjecciones o risas espontáneas de los participantes como respuesta a los estímulos visuales, pero de poca utilidad por la escasa información capturada. Los ficheros de audio se componen de dos canales capturados a 44.1 kHz: el primero (o canal izquierdo en una señal interpretada como estereofónica) contiene la señal de audio de los posibles ruidos ambientales y del sonido del vídeo usado como estímulo; el segundo canal (o canal derecho) contiene el audio generado por el usuario participante capturado a través de un micrófono acoplado a su cabeza.

4. Señales fisiológicas del sistema nervioso central, compuestas por 32 canales de EEG capturados con el casco *Biosemi active II system*[1] a una velocidad de 256 Hz, registradas mediante electrodos activos AgCl ubicados de acuerdo al sistema internacional 10-20 [24]. La disposición de los 32 electrodos a lo largo del casco se muestra en la figura 6.1.

5. Señales fisiológicas del sistema nervioso periférico: ECG, GSR, amplitud de la respiración y temperatura de la piel, capturadas con un dispositivo *Biosemi active II system* a una velocidad 1.024 Hz y muestreadas a 256 Hz. La información del ECG fue capturada mediante tres sensores colocados en el pecho del participante. Los datos sobre GSR fueron tomados con dos electrodos colocados en la falange distal de los dedos índice y corazón. La amplitud de la respiración mediante un cinturón de respiración colocado en el abdomen del participante. Por último, la temperatura de la piel se registró con la colocación de un sensor de temperatura en el dedo meñique.

Las señales fisiológicas se proporcionan de forma conjunta en 47 canales en el formato proporcionado por el dispositivo Biosemi (formato BDF, *Biosemi Data Format* en inglés), que es fácilmente legible con herramientas como MATLAB y EEGLAB[2]. La información contenida en los ficheros BDF se ajusta al formato

[1] http://www.biosemi.com
[2] http://sccn.ucsd.edu/eeglab

Índice de la etiqueta	**Etiqueta emocional**	**Número de muestras**
0	Neutral	111
1	Ira	14
2	Asco	56
3	Miedo	38
4	Alegría	88
5	Tristeza	69
6	Sorpresa	26
11	Diversión	100
12	Ansiedad	34
	TOTAL	**536**

Tabla 6.1: Emociones y estados recogidos en MAHNOB-HCI

descrito en la tabla 6.2. En las figuras 6.2 y 6.3 se muestran dos capturas de pantalla de las posibilidades para el tratamiento de las señales EEG ofrecidas por MATLAB y EEGLAB conjuntamente.

Todas las señales registradas en los ficheros BDF incluyen 30 segundos antes y 30 después de la visualización del estímulo y, según lo establecido por la metodología de grabación de Biosemi, no fueron grabadas tomando con referencia un electrodo concreto, por lo que para su tratamiento fue necesario realizar una referencia previa de las señales, como por ejemplo una referencia a la media.

Canal	**Descripción de la señal**	**Unidad física**
1-32	Señales capturadas por los 32 electrodos del EEG	microvoltios
33-35	Información del ECG	microvoltios
36-40	Sin uso	-
41	Información GSR	ohmios
42-44	Sin uso	-
45	Información de la amplitud de la respiración	microvoltios
46	Información de la temperatura de la piel	grados centígrados
47	Información de estado	valores enteros

Tabla 6.2: Estructura de los datos fisiológicos contenidos en los ficheros BDF de MAHNOB-HCI

6.2. Selección y extracción de características EEG

A los sujetos participantes en la creación del corpus se les pidió que tras la visualización del vídeo estímulo lo etiquetaran de acuerdo a su propia percepción.

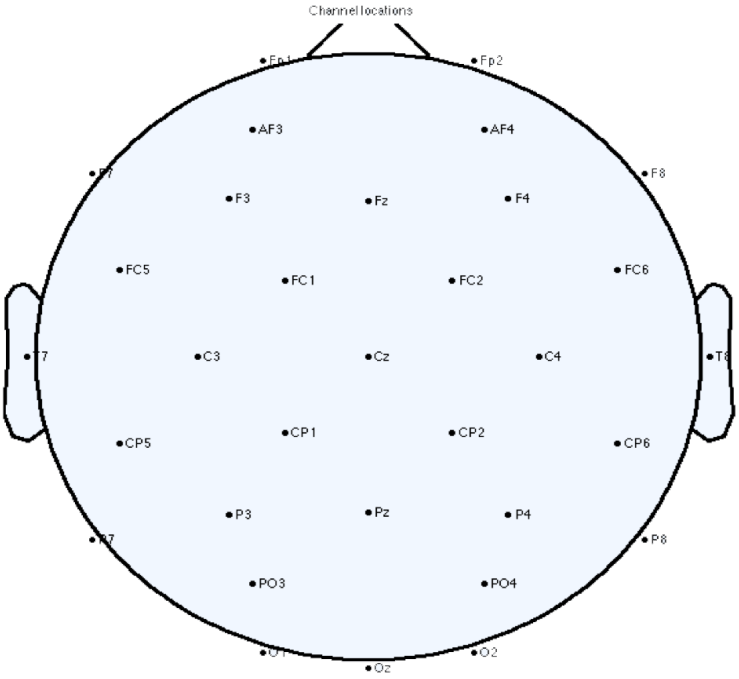


Figura 6.1: Posicionamiento de los 32 electrodos en el casco BIOSEMI según el estándar internacional 10-20

En concreto se les solicitó respuesta a cinco preguntas: etiqueta emocional y valores para la Activación, Valencia, Dominancia y predictibilidad.

Las posibles etiquetas emocionales podían pertenecer a una de las siguientes clases: Neutral, Ansiedad, Diversión, Tristeza, Alegría, Asco, Ira, Sorpresa y Miedo, codificadas mediante un valor entero.

Las otras cuatro preguntas se valoraron mediante una escala de valores enteros en el rango 1-9, correspondiendo para la Activación el valor 1 para una percepción de calma o pasividad y el valor 9 para una sensación de actividad máxima o excitación; para la Valencia un valor 1 para una percepción no placentera y un valor 9 para una completamente placentera; para la Dominancia un valor 1 para una sensación en la que no se tiene el control y un valor 9 para una sensación de pleno control; para la predictibilidad un valor 1 para la percepción de que la secuencia de eventos visionados fue impredecible o sorprendente y un valor 9 si fue completamente predecible.

Esta información reportada por los participantes se encuentra almacenada en ficheros XML para cada una de las sesiones grabadas. Para la selección de los datos a considerar se examinaron los ficheros *session.xml*, seleccionándose únicamente aquellos sujetos que tenían todos sus atributos afectivos completos y que, además,

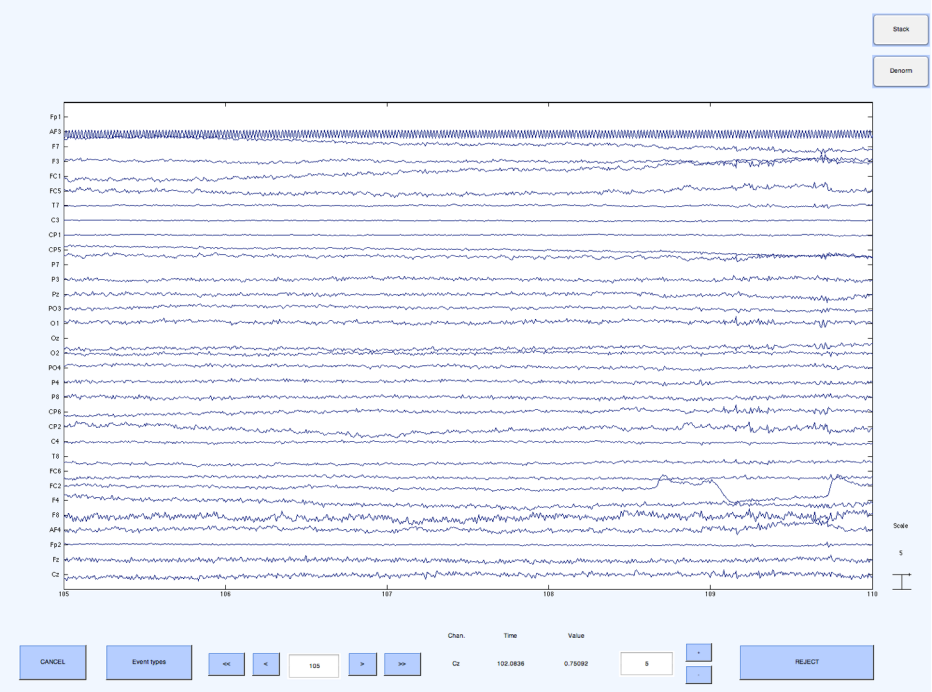


Figura 6.2: Ejemplo de una muestra de las señales correspondientes a los 32 electrodos del EEG procesados con MATLAB y EEGLAB

pertenecían al experimento de tipo *"emotion elicitation"*.

Para el tratamiento y posterior clasificación de las señales EEG, debido a la gran cantidad de información que una muestra obtenida a una velocidad de 256 Hz contiene, aspecto que haría intratable su posterior procesado, se procedió a la obtención de unas características basadas en la energía que describieran los datos consistentemente. Para ello se obtuvo el cálculo de los logaritmos de la potencia espectral para las bandas de frecuencia θ (entre 4 y 8 Hz), α (entre 8 y 12 Hz), α lenta (entre 8 y 10 Hz), β (entre 12 y 30 Hz) y γ (superiores a 30 Hz), para cada uno de los 32 electrodos. Las ondas δ, con rango de frecuencia inferior a los 4 Hz no se consideraron por estar relacionadas con estados de sueño profundo no relevantes para el estudio.

De modo similar, también se obtuvieron características basadas en la diferencia de potencia espectral para cada una de las 14 parejas simétricas de electrodos ubicados en cada hemisferio en las bandas θ, α, β y γ, con el objetivo de medir posibles simetrías o asimetrías en la actividad cerebral durante la percepción de los diferentes estímulos [174, 48].

En total se obtuvieron 216 características basadas en la potencia espectral de las señales EEG. Estas características son algunas de las habitualmente empleadas en el análisis de señales EEG [169, 112, 101, 194].

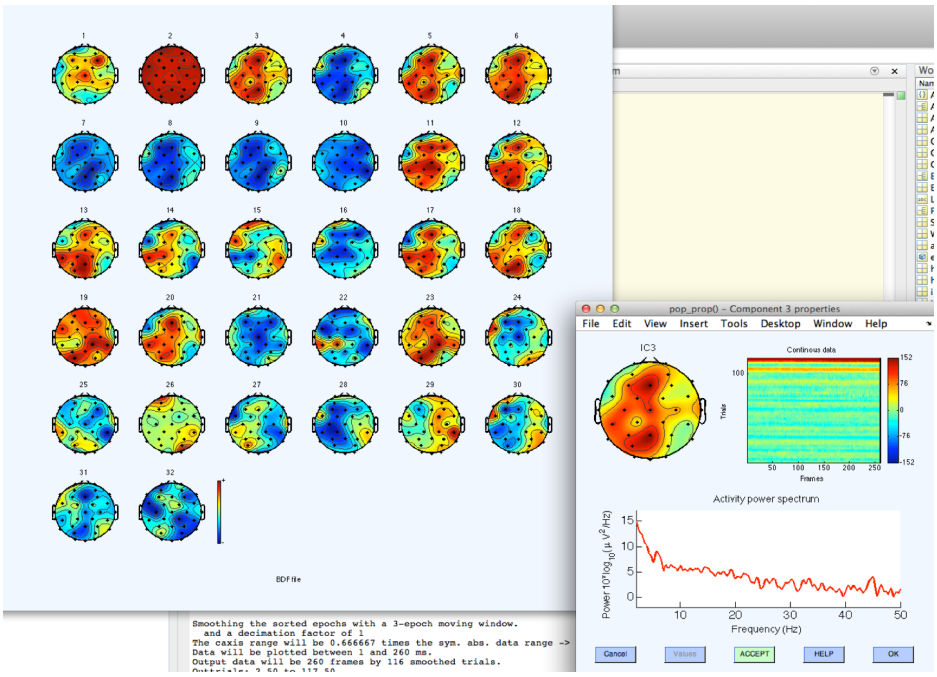


Figura 6.3: Ejemplo de visualización de la actividad de cada uno de los 32 electrodos del EEG como valores del logaritmo de la potencia espectral y de la frecuencia

6.3. Clasificación y análisis de resultados

A partir del conjunto de muestras y emociones detallado en la tabla 6.1 y de las características obtenidas para las señales EEG descritas en el apartado anterior, en una primera aproximación se evaluaron diferentes esquemas de clasificación para las 9 clases de emociones disponibles en la base de datos. Para ello, con el objetivo de reducir posibles diferencias entre sujetos, las características de cada muestra fueron normalizadas en el rango de valores [0, 1], substrayendo el valor mínimo del conjunto de valores de una característica dada para un sujeto determinado y dividiéndola por la diferencia de sus valores máximo y mínimo.

No se obtuvieron resultados relevantes en los métodos de clasificación aplicados: SVM, K-NN, LMT, C4.5 o perceptrón multicapa, entre otros, debido a que mientras algunas de las 9 emociones eran clasificadas con gran precisión, otras presentaban un rendimiento inferior al de un clasificador simplemente aleatorio. En la tabla 6.3 se muestra un resumen de algunos de los algoritmos de clasificación aplicados y la cobertura media alcanzada en el reconocimiento de las 9 emociones recogidas en MAHNOB-HCI. En ella se observa que el algoritmo que presentó una tasa de acierto mayor fue LMT, con una cobertura media de un 22,95 % (valor-F = 0,19), que supone una tasa de acierto del doble que la obtenida con un clasificador completamente aleatorio. Adicionalmente, de las matrices de confusión obtenidas

por cada uno de los clasificadores, las emociones que mejor ratio de clasificación media presentaron fueron las correspondientes a las emociones Neutral, Alegría y Diversión, que coinciden con las clases que mayor número de muestras disponen. Las tasas de de acierto para estas tres emociones frente al resto (valor-F entre paréntesis) se muestran en la tabla 6.4.

El entrenamiento y la evaluación de los clasificadores se realizó mediante la aplicación de una estrategia de validación cruzada sobre el conjunto total de 536 muestras, utilizando en cada una de las 536 iteraciones de entrenamiento / validación una única muestra como test y el resto como conjunto de entrenamiento (estrategia *leave-one-out cross-validation*), obteniendo la precisión de la clasificación como la media aritmética de cada una de las iteraciones.

Clasificador	**Cobertura media**	**valor-F**
SVM	22,20 %	0,16
K-NN	20,15 %	0,18
LMT	22,95 %	0,19
C4.5	19,60 %	0,19
Perceptrón multicapa	17,54 %	0,17
Aleatorio	11,11 %	0,11

Tabla 6.3: Resultados obtenidos mediante diferentes estrategias de clasificación para las 9 emociones recogidas en MAHNOB-HCI

Clasificador	**Neutral**	**Alegría**	**Diversión**
SVM	64,90 % (F=0,37)	20,50 % (F=0,23)	28,00 % (F=0,23)
K-NN	47,70 % (F=0,34)	20,50 % (F=0,20)	21,00 % (F=0,22)
LMT	58,60 % (F=0,39)	22,70 % (F=0,22)	27,00 % (F=0,24)
C4.5	39,60 % (F=0,28)	11,40 % (F=0,27)	15,00 % (F=0,18)
Perceptrón multicapa	26,70 % (F=0,32)	16,10 % (F=0,13)	23,10 % (F=0,18)

Tabla 6.4: Cobertura media obtenida (valor-F entre paréntesis) en la clasificación de las emociones Neutral, Alegría y Diversión para MAHNOB-HCI

Alternativamente, se realizó una selección de características mediante un test ANOVA simple tomando como variable independiente la clase y descartando las no significativas con $p > 0,05$, quedando como resultado del test únicamente 16 características. Sobre estas características se aplicaron los mismos algoritmos de clasificación para la detección de las 9 emociones que en el caso anterior con similares resultados. En la tabla 6.5 se muestran los resultados de la clasificación sobre la selección de características anteriormente descritas. En este caso, el clasificador que presentó un rendimiento ligeramente superior fue el Perceptrón multicapa, con una cobertura media de un 24,30 % (valor-F = 0,19). En cuanto a las emociones clasificadas con mayor tasa de acierto, al igual que en el caso anterior, Neutral, Alegría y Diversión fueron las emociones que destacaron frente a las seis restantes.

Clasificador	Cobertura media	valor-F
SVM	23,51 %	0,14
K-NN	19,40 %	0,18
LMT	24,10 %	0,17
C4.5	17,54 %	0,17
Perceptrón multicapa	24,30 %	0,19
Aleatorio	11,11 %	0,11

Tabla 6.5: Resultados en la clasificación de las 9 emociones recogidas en MAHNOB-HCI tras realizar una selección de características mediante un test ANOVA

A la vista de los resultados mostrados en las tablas 6.3, y 6.5 y de modo equivalente a como se realizó en la experimentación llevada a cabo por los creadores de MAHNOB-HCI [169], se decidió agrupar las 9 emociones sobre dos dimensiones afectivas: Activación y Valencia, con el objetivo de evaluar diferentes esquemas de clasificación. La distribución de las emociones en cada una de estas clases se realizó según se describe en las tablas 6.6 y 6.7.

Clases de Activación	Emociones	Número de Muestras
Relajación	Neutral, Asco, Tristeza	236
Media	Alegría, Diversión	188
Excitación	Ira, Miedo, Sorpresa, Ansiedad	112
TOTAL		**536**

Tabla 6.6: Distribución de las 9 emociones de MAHNOB-HCI en tres clases de Activación

Clases de Valencia	Emociones	Número de Muestras
Negativa	Ira, Asco, Miedo, Tristeza, Ansiedad	211
Neutral	Neutral, Sorpresa	137
Positiva	Alegría, Diversión	188
TOTAL		**536**

Tabla 6.7: Distribución de las 9 emociones de MAHNOB-HCI en tres clases de Valencia

Para la clasificación de las muestras en las tres clases de Activación (Relajación, Media y Excitación) y las tres de Valencia (Negativa, Neutral y Positiva) se aplicaron los mismos esquemas de clasificación que los empleados anteriormente en la predicción del conjunto total de emociones: SVM, K-NN, LMT, C4.5 y perceptrón multicapa. En la tabla 6.8 se muestra el rendimiento de cada clasificador para ambas dimensiones afectivas. Todas las experimentaciones fueron realizadas siguiendo una estrategia de tipo *test leave-one out cross-validation*. En este caso,

los dos métodos de clasificación que reportaron mejores resultados fueron SVM y LMT, con tasas de aciertos medias del 48 % para la Activación y del 47 % para la Valencia.

Clasificador	**Activación**	**Valencia**
SVM	48,13 % (F=0,46)	47,20 % (F=0,47)
K-NN	45,15 % (F=0,40)	39,40 % (F=0,39)
LMT	47,95 % (F=0,47)	47,01 % (F=0,46)
C4.5	40,70 % (F=0,40)	41,04 % (F=0,40)
Perceptrón multicapa	41,04 % (F=0,40)	38,00 % (F=0,38)
Aleatorio	33,33 % (F=0,33)	33,33 % (F=0,33)

Tabla 6.8: Resultados de la cobertura media en la clasificación de las 9 emociones de MAHNOB-HCI sobre las dimensiones de Activación y Valencia

Del mismo modo a como se procedió anteriormente, sobre las muestras agrupadas en las dos dimensiones afectivas se realizó una selección de características mediante un test ANOVA simple, tomando como variable independiente la clase y descartando las no significativas con $p > 0,05$, quedando como resultado del test únicamente 27 características para la Activación y 21 para la Valencia. Sobre estas características se aplicaron los mismos algoritmos de clasificación con los resultados que se muestran en la tabla 6.9. En ella se observa que las coberturas medias alcanzadas son similares a las obtenidas previamente sin ningún tipo de selección de características, lo que contrasta con los resultados obtenidos por los creadores de la base de datos [169], donde se reportan tasas medias de aciertos de un 52,4 % (F=0.42) y de un 57,0 % (F=0,56) para la Activación y la Valencia, respectivamente.

Clasificador	**Activación**	**Valencia**
SVM	46,30 % (F=0,43)	47,40 % (F=0,46)
K-NN	45,15 % (F=0,45)	42,53 % (F=0,41)
LMT	48,32 % (F=0,43)	47,40 % (F=0,45)
C4.5	41,05 % (F=0,40)	42,53 % (F=0,40)
Perceptrón multicapa	45,52 % (F=0,42)	43,10 % (F=0,37)
Aleatorio	33,33 % (F=0,33)	33,33 % (F=0,33)

Tabla 6.9: Resultados de la cobertura media en la clasificación de las 9 emociones de MAHNOB-HCI sobre las dimensiones afectivas de Activación y Valencia tras realizar una selección de características mediante un test ANOVA

Como alternativa, se buscó un enfoque de clasificación diferente basado en la cuantificación de las emociones sobre las dimensiones de Valencia y Activación, similar al trabajo realizado por Johnny Fontaine [70]. Este estudio, enmarcado dentro del proyecto GRID[3] de investigación internacional en el ámbito de las ciencias afectivas, establece un modelo multidimensional e intercultural para un

[3]http://www.affective-sciences.org/grid

total de 24 emociones. El modelo describe 4 dimensiones (Valencia, Potencia, Activación y Novedad) y se fundamenta en un total de 142 características emotivas basadas, a su vez, en 5 componentes emocionales que preparan al organismo a reaccionar adaptativamente ante un evento. Estos componentes son: 1) Estimación de las emociones: 2) Reacciones corporales; 3) Expresiones (vocales, faciales y gestuales); 4) Tendencias de acción; y 5) Sentimientos. Las características fueron valoradas mediante cuestionarios GRID traducidos a más de 20 idiomas, mediante los cuales se les solicitó a los sujetos participantes que asignaran en una escala de 1 a 9 en qué medida cada característica emotiva podría ajustarse a una emoción particular en su contexto lingüístico y cultural. La escala de valores usados en el cuestionario fue: a) Completamente improbable: 1-3; b) Ni improbable, ni probable: 4-7; c) Altamente probable: 8-9.

A través de los datos recogidos mediante los cuestionarios GRID y mediante técnicas de análisis factorial con rotación VARIMAX, se establecieron las relaciones entre las dimensiones de Valencia, Potencia, Activación y Novedad, cuantificándose numéricamente las emociones sobre cada una de estas dimensiones [70]. En la figura 6.4 se presenta la relación entre la dimensión de la Valencia y la Activación, mientras que en la tabla 6.10 se relacionan los valores medios en ambas dimensiones para las 9 emociones recogidas en MAHNOB-HCI. A la tabla original de valores medios se le ha añadido la expresión Neutral y se ha asimilado la emoción Diversión de MAHNOB-HCI a la emoción Placer del proyecto GRID.

Emoción	**Valencia**	**Activación**
Ira	-1,03	0,83
Ansiedad	-0,45	1,24
Asco	-0,98	-0,43
Miedo	-0,46	1,25
Alegría	1,54	0,27
Neutral	0,00	0,00
Placer (Diversión)	1,53	0,05
Tristeza	-0,51	-1,46
Sorpresa	0,18	0,55

Tabla 6.10: Valores medios obtenidos mediante análisis factorial con rotación VARIMAX en las dimensiones de Valencia y Activación y adaptación para las 9 emociones recogidas en MAHNOB-HCI

Con el conjunto de valores medios definidos para cada emoción y dimensión y después de evaluar el rendimiento de diversos esquemas de clasificación valorando aquellos que minimizaban la raíz del error cuadrático medio ($RMSE$), se implementaron dos clasificadores de tipo SVR con kernel radial (RBF con parámetros $C = 10$ y $G = 0,01$ y error $RMSE = 0,96$) para intentar clasificar las señales EEG en las dimensiones de Valencia y Activación. El propósito de estos clasificadores fue predecir el valor numérico de la emoción sobre cada dimensión afectiva (valores del eje de abscisas para la Valencia y del eje de ordenadas para la Activación de la figura 6.4), para posteriormente combinar sus resultados de modo

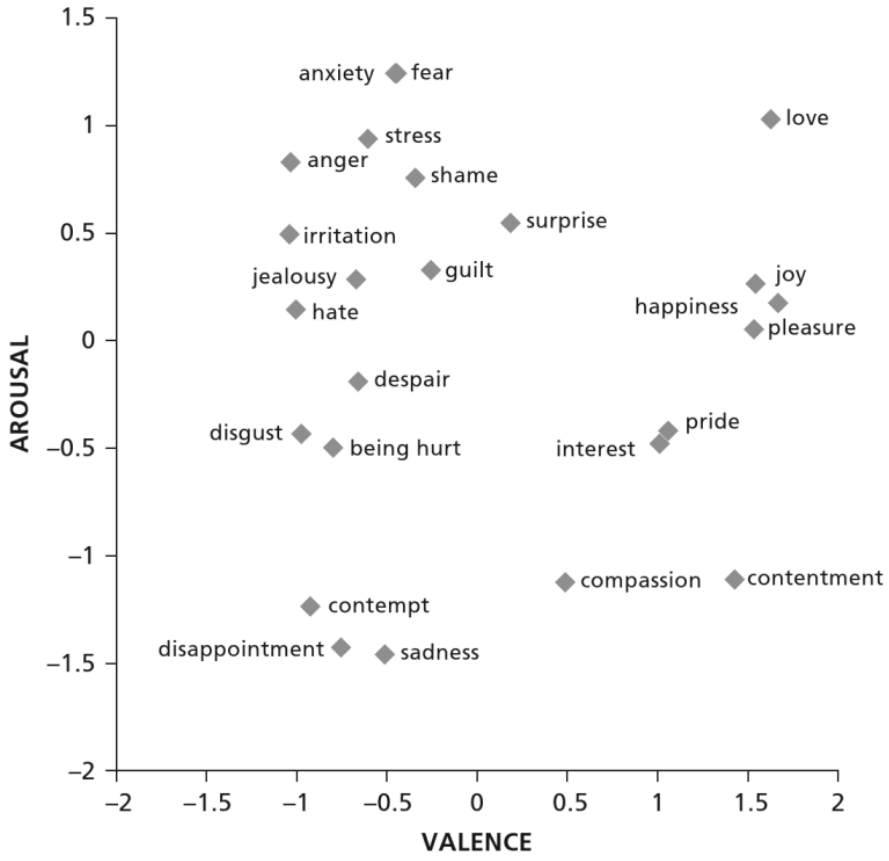


Figura 6.4: Representación de las 24 emociones recogidas en el proyecto GRID sobre las dimensiones de la Valencia y la Activación (*fuente Johnny Fontaine 2013*)

que se pudiera predecir la emoción real en función de los valores proporcionados por el proyecto GRID detallados anteriormente en la tabla 6.10.

Los clasificadores implementados (SVM, K-NN, LMT, J48 y Percetrón multicapa) no pudieron determinar con precisión las emociones subyacentes a partir de las señales EEG contenidas en MAHNOB-HCI. En la figura 6.5 se visualizan las 536 muestras EEG por emoción tras la predicción realizada con la SVR sobre los valores GRID de cada muestra sobre las dimensiones de Valencia y Activación. En el gráfico se observa que la mayor parte de las muestras se encontraban concentradas en torno al eje de coordenadas (emoción Neutral). El resto de las muestras que aparecen dispersas por el gráfico se encuentran entremezcladas, motivo por el que no posible clasificar las señales EEG mediante la aproximación basada en la cuantificación de las emociones sobre las dimensiones de Valencia y Activación, según los datos proporcionados por el proyecto GRID.

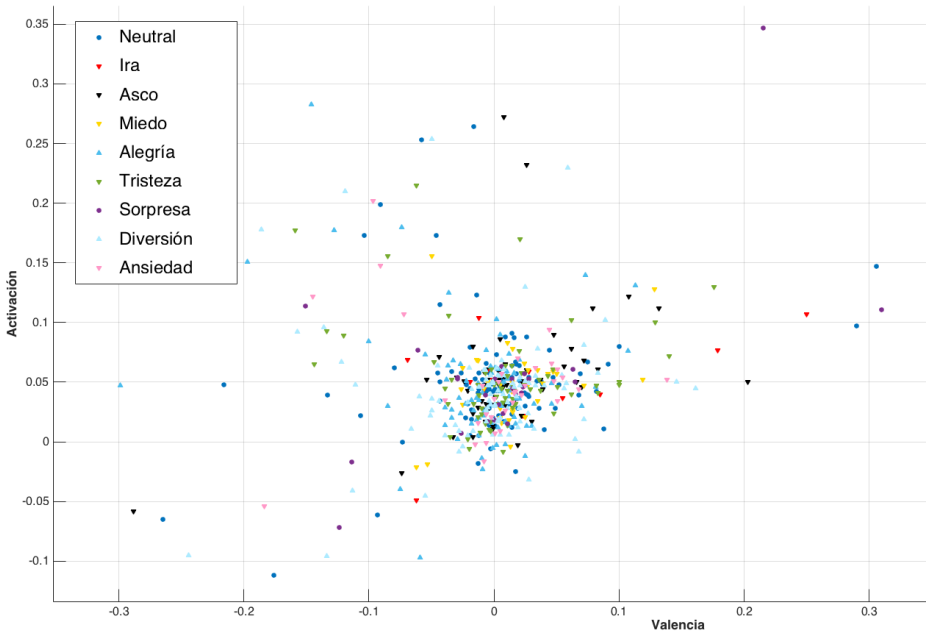


Figura 6.5: Distribución de las muestras EEG de MAHNOB-HCI cuantificadas con los valores del proyecto GRID y clasificadas mediante SVR

6.4. Conclusiones

Las señales fisiológicas, así como la combinación de varias de ellas, pueden aportar valiosa información sobre el estado afectivo del sujeto. Sin embargo, la alta variación de estas señales entre individuos hace que resulten complejas de analizar y clasificar mediante patrones que puedan definir de forma precisa el estado afectivo de la persona. Tomando en consideración que los datos incluidos en MAHNOB-HCI fueron recogidos en un entorno controlado de laboratorio y que las diferentes estrategias de clasificación probadas sobre las señales EEG no fueron capaces de obtener una alta precisión, es de esperar que en un entorno no controlado, donde probablemente la información capturada sea de peor calidad, la detección de patrones afectivos se torne más compleja e imprecisa.

Por otro lado, la obtención de gran parte de las señales fisiológicas analizadas en este apartado requiere para su captura dispositivos costosos, precisos y de difícil calibración, pero sobre todo altamente intrusivos para el usuario, lo que supone que en ciertos entornos no resulte posible ni adecuado su uso. No obstante, recientemente están emergiendo prendas y objetos de uso cotidiano (*wearables* en inglés) capaces de monitorizar algunos parámetros de la actividad fisiológica de una persona de un modo muy poco intrusivo. Objetos como pulseras o relojes que miden el pulso cardíaco, la temperatura o la presión sanguínea o camisetas que incorporan sensores en su cuerpo y que se integran de forma inalámbrica con

teléfonos inteligentes u ordenadores personales, pueden abrir nuevos horizontes en el campo de la computación afectiva mediante el análisis de características fisiológicas de larga duración que permitan la detección de patrones y, por ende, de ciertos estados afectivos del sujeto que los viste.

Parte III

Aportaciones en la mejora instruccional

Los STI son ampliamente usados en contextos educativos [77, 66]. Uno de sus retos se encuentra relacionado con el nivel de ayuda más adecuado que debe ser proporcionado al estudiante, no sólo para conseguir sus metas, sino también para mantenerlo en un estado afectivo que le permita mejorar su aprendizaje según los objetivos propuestos por el STI [52].

En este contexto, en el siguiente capítulo se detallará la aportación realizada en el ámbito de la computación afectiva al dotar a un STI con la capacidad para la valoración del estado afectivo del usuario, con el objetivo de proporcionar al sistema información adicional para la decisión del nivel de ayuda más adecuado a proveer al alumno durante su interacción con el mismo.

El objetivo es demostrar la hipótesis de que es posible intervenir en el estado afectivo del usuario mediante la aplicación de técnicas instruccionales que tomen en consideración las valoraciones del estudiante sobre las dimensiones de la Dominancia, la Valencia y la Activación, ajustando la respuesta del STI de modo que ésta influya en su estado afectivo y, previsiblemente, en su rendimiento final.

Capítulo 7

Incorporación de soporte afectivo a un STI

Resumen
En este capítulo se describirá el método seguido para dotar de soporte afectivo a un STI para el aprendizaje de la aritmética y el álgebra lineal, denominado HBPS, mediante una aproximación de tipo *sensor-free*, con el propósito de ajustar la respuesta del STI de modo que ésta pueda influir en el estado afectivo del estudiante.

Contenidos

7.1. Fundamentos de HBPS 103
7.2. Diseño del núcleo afectivo del STI 105
7.3. Construcción del núcleo afectivo del STI 107
7.4. Implementación y análisis de los clasificadores afectivos . . . 115
7.5. Experimentación 117
7.6. Análisis de resultados 119
7.7. Conclusiones y Discusión 121

Uno de los grandes retos de los STI se encuentra relacionado con el nivel de ayuda más adecuado que puede ser proporcionado al estudiante con el objetivo de mejorar su motivación y, en consecuencia, su aprendizaje final [52]. En este sentido, existe diversidad de criterios a favor y en contra de las técnicas instruccionales utilizadas para guiar al alumno de forma progresiva hacia la mejora del aprendizaje [79, 13].

Sin embargo, aunque en diversos estudios se reportan sólidas evidencias de que el estado emocional del estudiante puede tener un impacto significativo en su motivación y, por tanto, en su rendimiento [3, 142], la mayoría de los STI

existentes no han tomado en consideración este hecho, concentrando sus esfuerzos exclusivamente en la maximización del conocimiento adquirido por el estudiante, sin valorar el impacto que pueden suponer las decisiones de carácter instruccional en las variables afectivas del usuario. No obstante, en los últimos años ha habido un interés creciente en esta dirección con la incorporación de mecanismos de soporte afectivo de baja intrusividad [192, 32, 9, 55], analizando señales procedentes de la voz [122, 80], la postura [50, 74] o de expresiones faciales [170, 42] o, incluso, mediante otros mecanismos más intrusivos que utilizan información fisiológica procedente de señales EEG [170, 6, 96].

Con el objetivo de poder demostrar la hipótesis de que es posible ajustar la respuesta del STI para que ésta influya en el estado afectivo del usuario, a lo largo de este capítulo se presentará la extensión realizada sobre HBPS (solucionador de problemas basado en hipergrafos o *Hypergraph Based Problem Solver* en inglés) [7] mediante el diseño e implementación de un módulo para la valoración del estado afectivo del estudiante en función del nivel de ayuda proporcionado por el sistema. Con esta extensión se pretende dotar al STI de capacidad para la determinación del nivel más apropiado de ayuda a proveer al estudiante según su modelo cognitivo, el contexto formativo y su estado afectivo subyacente.

Para la valoración del estado afectivo del usuario, se consideraron las variables afectivas correspondientes a las dimensiones del modelo emocional propuesto por Russell y Mehrabian [153, 127]. Este modelo, conocido comúnmente como modelo PAD (acrónimo del inglés *Pleasure-Arousal-Dominance*), ha sido extensivamente utilizado en la literatura relacionada con la detección del estado afectivo de un usuario [97, 134]. El modelo sitúa todas las posibles emociones en un espacio tridimensional definido por la Valencia, la Activación y la Dominancia. Los valores de cada dimensión pueden ser medidos mediante la Escala Diferencial Semántica (*Semantic Differential Scale* en inglés) [128] o con el uso de test SAM (*Self-Assessment Manikin* en inglés) [27]. La Escala Semántica requiere realizar 18 valoraciones diferentes en una escala de 9 puntos, relacionadas con el modelo PAD mediante un análisis de factores. Por el contrario, los test SAM tan solo requieren tres valoraciones, también en una escala de 9 puntos, que pueden ser directamente relacionadas con las tres dimensiones definidas en el modelo PAD, aspecto que hace su uso más práctico.

El objetivo perseguido es intentar mantener al usuario en un estado afectivo que favorezca su interés en los contenidos pedagógicos propuestos, determinando para ello el nivel de ayuda más adecuado a proporcionar en cada momento, con el propósito de influir en su estado afectivo y, de este modo, poder mejorar su rendimiento.

El módulo se entrenó y se validó con la información recogida mediante experimentaciones previas realizadas en entornos educativos reales con HBPS, recabando información del estado afectivo del usuario en las dimensiones de Dominancia, Activación y Valencia mediante el uso de test SAM.

En los siguientes apartados se introducirá, en primer lugar, los fundamentos de HBPS para, posteriormente, formular y detallar la aproximación seguida para dotar al STI de soporte afectivo para la determinación de las ayudas instruccionales a proporcionar al estudiante durante su interacción con el sistema.

7.1. Fundamentos de HBPS

HBPS [7] es un STI especialmente diseñado para paliar las dificultades con las que frecuentemente suelen encontrarse los estudiantes cuando se inician en el aprendizaje de problemas matemáticos descritos mediante enunciados y que, básicamente, consisten en su traslación a una notación simbólica y su resolución final.

El núcleo de HBPS se basa en la implementación de un motor de inferencia y un mecanismo de representación del conocimiento, ambos fundamentados en un lenguaje descriptivo basado en hipergrafos, así como en la idea del uso de esquemas conceptuales para representar el conocimiento del estudiante. Su modelo de usuario alberga información sobre el nivel de competencia en el uso de varios esquemas conceptuales [121, 150], así como en la aplicación del método cartesiano [145]. Esta información es usada por el STI para la adaptación de los mensajes de ayuda instruccional facilitados en su interacción con el sistema en función de su estado actual de conocimiento y del problema que se está resolviendo.

El enfoque de tutorización adoptado en HBPS se basa en una aproximación no guiada, proporcionando al estudiante una mayor autonomía al permitirle definir las cantidades que necesite en cada momento del desarrollo de la solución al problema planteado, sin la imposición de ningún tipo de restricción en la búsqueda del camino de resolución.

El propósito instruccional de HBPS es ayudar al estudiante durante la resolución de problemas matemáticos propuestos mediante enunciados verbales, siguiendo un enfoque algebraico o aritmético (modo de trabajo configurable en el STI), tomando como modelo de referencia resolutivo el método cartesiano [67, 145]. De este modo, el problema puede ser descrito por una secuencia de acciones ordenadas: 1) la asignación de n letras para la representación de las n cantidades desconocidas; 2) la definición del resto de cantidades desconocidas usando expresiones algebraicas o aritméticas cuando la estructura del problema lo permite; 3) el establecimiento de un conjunto de n ecuaciones, partiendo de la base de que las n cantidades pueden ser expresadas de formas diferentes; y 4) la resolución del conjunto de las n ecuaciones.

HBPS permite al estudiante completa libertad en cuanto a la definición del número de variables a definir, así como el camino de resolución escogido. No obstante, el diseño de su interfaz gráfico de usuario (GUI o *Graphical User Interface* en inglés) impone tres restricciones basadas en la aplicación del método cartesiano en lo que se refiere a las posibles acciones que el usuario puede realizar: 1) asignar una variable a una cantidad; 2) expresar una cantidad desconocida en términos de otras previamente definidas; y 3) definir ecuaciones relacionando varias cantidades previamente definidas. Estas acciones son estratégicamente secuenciadas en HBPS para fomentar una sistemática, estructurada y organizada resolución del problema utilizando el método cartesiano.

Otras restricciones básicas que impone HBPS son: la necesidad de definir las cantidades desconocidas antes de que puedan ser usadas en expresiones matemáticas; y que las expresiones se limiten a relaciones ternarias, es decir a relaciones de

sumas o productos de términos, permitiendo al estudiante simplificar relaciones complejas en otras más triviales.

El diseño del GUI refleja estas intenciones metodológicas. En la figura 7.1 se muestra una captura del interfaz gráfico de HBPS en un momento concreto de la resolución de un problema. La parte inferior de esta pantalla se encuentra dividida en dos secciones diferenciadas. La sección de la izquierda es la destinada a la definición de nuevas cantidades mediante la introducción de una nueva letra o de una expresión relacionada con cantidades anteriormente definidas. En el caso de que el estudiante defina la nueva cantidad mediante la asignación de una letra, también deberá asignar a dicha letra el significado de la variable, seleccionando éste entre una lista de posibles valores. Estas posibles descripciones son extraídas de una base de conocimiento, permitiendo al componente de dominio asignar unívocamente la letra de la variable con la cantidad correcta.

Como se observa en la figura 7.1, el estudiante va construyendo las expresiones usando un componente gráfico similar al de una calculadora, con botones para cada uno de los posibles operadores aritméticos a utilizar, así como otros botones para cada una de las cantidades previamente definidas por el mismo. En el caso de que la expresión definida sea correcta, se infiere la cantidad que la representa y se asigna automáticamente una descripción textual a la expresión.

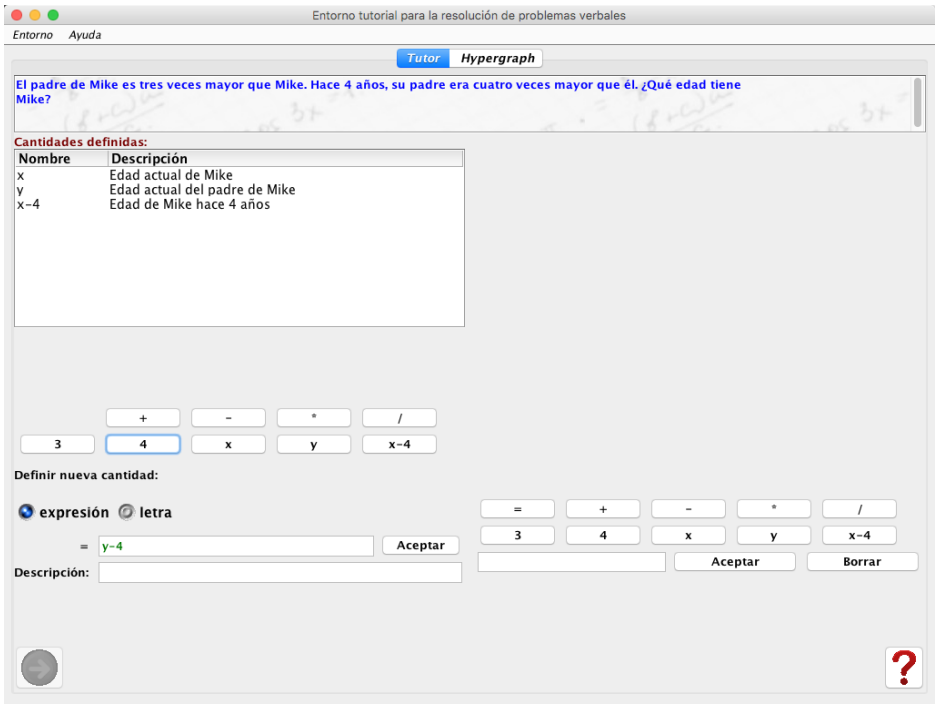


Figura 7.1: Interfaz gráfico de HBPS

La sección de la derecha de la pantalla es donde el alumno puede construir ecuaciones mediante el uso de un sistema de botones similar, también en este

caso, al de una calculadora. Sus botones contienen los mismos elementos que los del panel de la izquierda, a excepción de la incorporación del botón de igualdad utilizado, en este bloque, para la asignación de cantidades o expresiones.

Con el fin de ayudar a identificar visualmente las cantidades definidas éstas se muestran en el interior de una tabla situada en la parte superior del interfaz. En ella se incluye el nombre asignado y su descripción asociada. De modo similar, cada vez que el alumno define una nueva cantidad, ésta se añade a la tabla y se crea un nuevo botón en ambos paneles que le permitirá usar la nueva variable para definir otras adicionales o para usarla en la construcción de una ecuación como parte del camino de resolución del problema. Adicionalmente, como medio de ayuda al estudiante, se visualiza un mensaje emergente con la descripción de la cantidad cada vez que el puntero del ratón se posiciona sobre el botón que la identifica.

Las ayudas instruccionales adaptativas provistas por el HBPS se basan en la información que se tiene de la competencia del estudiante en el uso de algún esquema conceptual fundamental [121, 150]. Estos esquemas fundamentales se sustentan en la comprensión de ciertos tipos de relaciones entre cantidades. En general, los enunciados de los problemas raramente revelan estas relaciones de forma explícita, sino que, por el contrario, deben inferirse desde los esquemas conceptuales subyacentes en la situación descrita en el enunciado. Investigaciones consolidadas sugieren que el éxito en la resolución de un problema se apoya fuertemente en los esquemas conceptuales a los que una persona puede recurrir. Por este motivo, la capacidad del estudiante en el uso de cada esquema conceptual proporciona un gran nivel de conocimiento que puede ser explotado con el objetivo de proveer ayudas personalizadas.

7.2. Diseño del núcleo afectivo del STI

A diferencia de otros trabajos [45, 53, 41, 155, 141, 140], el diseño del núcleo para la valoración del estado afectivo del estudiante no se orientó a la construcción de un detector emocional. La intención fue la creación de un módulo para la valoración del estado afectivo del usuario en función de la ayuda provista al alumno sobre las dimensiones afectivas de la Dominancia, la Valencia y la Activación, para ser integrado en HBPS y prestarle soporte en la decisión de la ayuda instruccional más adecuada a proporcionar al estudiante durante el proceso de resolución de un problema, de acuerdo a la dimensión afectiva que en cada momento se desee mejorar.

Dado que HBPS proporciona ayudas instruccionales personalizadas cada vez que el sistema detecta que una intervención es requerida, ya sea de forma explícita por el usuario o por un error cometido durante la resolución de un problema, no se consideró conveniente evaluar el impacto de un nivel concreto de ayuda con información basada exclusivamente en la siguiente interacción del usuario con el sistema, debido, principalmente, a los limitados efectos que una decisión tendría sobre una acción individual. Por este motivo, el STI se configuró para que el nivel de ayuda a proporcionar al alumno según su modelo de usuario, la dificultad del

problema a resolver y su estado afectivo, se fijara al inicio de la carga del problema y se mantuviera constante durante todo el proceso de resolución. De esta forma resultaría posible evaluar el impacto del nivel de ayuda proporcionado mediante la comparación de su estado previo antes de iniciar el problema y su estado final tras la resolución del mismo.

El módulo de valoración del nivel de ayuda óptimo se creó con el propósito de mejorar las dimensiones afectivas del alumno mediante la aplicación de dos estrategias de maximización M_i diferenciadas: Dominancia (M_1); Valencia y Activación conjuntamente (M_2).

La elección de estas estrategias no fue arbitraria; ambas se encuentran relacionadas con potenciales ganancias en el aprendizaje. Por un lado, los resultados reportados en [27] demuestran que la dimensión de la Dominancia explica la mayor parte de la varianza del par "guiado-autónomo" de la Escala Diferencial Semántica. Esto significa que una estrategia que persiga la maximización de la Dominancia contribuirá al incremento de la autonomía del estudiante. Por otro lado, el incremento individual de la Valencia (el placer) o la Activación (la excitación) no necesariamente conduce a ganancias en el aprendizaje, mientras que su incremento simultáneo se asocia con un mayor interés, motivación y compromiso.

A continuación, se formulará el problema de la determinación del nivel de ayuda más adecuado a proporcionar al estudiante para una determinada ayuda y contexto específico.

Para ello, consideremos un conjunto A de todas las posibles acciones instruccionales en un contexto concreto y una estrategia de maximización específica M. El nivel de consecución L_M tras la realización de la ayuda instruccional $a \in A$ puede expresarse como

$$L_M = f_M(a, contexto) \tag{7.1}$$

donde f_M es la función que se debe aprender y *contexto* el conjunto completo de hechos que pueden afectar al valor de L_M *a posteriori*, como puede ser la dificultad del problema a resolver, el modelo de conocimiento del usuario, su estado afectivo, etc. En la práctica el contexto suele modelarse como un vector d-dimensional $\{c_1, c_2, \ldots, c_d\}$, convenientemente escogido para alcanzar un balance entre la dimensionalidad de los datos y la suficiente precisión para la representación de los hechos más relevantes. En este sentido, L_M se aproximará como

$$L_M \approx f'_M(a, c_1, c_2, \ldots, c_d) \tag{7.2}$$

Una forma para conseguir aprender la función f_M es mediante el uso de técnicas de aprendizaje automático, utilizando, para ello, un número de muestras de ejemplo convenientemente etiquetadas de la forma $\{a, c_1, c_2, \ldots, c_d\} \rightarrow f_M$, que relacionan los parámetros de la función $\{a, c_1, c_2, \ldots, c_d\}$ con una etiqueta o valor que representa el nivel de consecución L_M tras la aplicación de la estrategia de maximización M. Estas muestras etiquetadas pueden obtenerse mediante el desarrollo de sesiones experimentales diseñadas con el objetivo de la recopilación de datos de entrenamiento. De este modo, el problema de estimar la función f'_M a partir de las muestras de entrenamiento puede ser realizado mediante métodos

tradicionales de clasificación o de regresión, dependiendo de si el objetivo de la variable L_M es categórico o continuo, respectivamente.

Una vez la función f'_M ha sido entrenada, la ayuda a que maximiza el nivel de consecución de una determinada estrategia de maximización M puede obtenerse como

$$\arg\max_{a \in A} f_M(a, c_1, c_2, \ldots, c_d) \tag{7.3}$$

7.3. Construcción del núcleo afectivo del STI

Para la construcción del núcleo afectivo del STI e implementación de los clasificadores que permitieran maximizar cada una de las estrategias M_i definidas anteriormente, se llevó a cabo una experimentación previa con el propósito de recabar un conjunto suficiente de datos que sirvieran para crear un sistema de aprendizaje automático que proporcionara la información afectiva necesaria para cada estrategia de maximización.

Para la experimentación se contó con una muestra de 48 estudiantes de quinto curso de educación Primaria, con edades comprendidas entre los 10 y los 11 años. Para su ejecución se diseñaron 10 problemas aritméticos con diferente grado de dificultad. Estos problemas se plantearon secuencialmente y en el mismo orden a todos los estudiantes. Para esta experimentación, HBPS se configuró para operar con problemas aritméticos y para que estableciera un nivel de ayuda inicial aleatorio para cada uno de los 10 problemas propuestos y para cada estudiante. Se definieron 4 niveles de ayuda con diferente nivel de detalle, siendo el nivel 1 la ayuda mínima facilitada para la acción en la que se encuentra incurso el alumno y el nivel 4 la más descriptiva. No obstante, el estudiante dispuso de completa libertad para solicitar al STI una nueva ayuda con un mayor nivel de detalle, hasta alcanzar la ayuda máxima que proporciona una completa descripción de la acción.

Este escenario experimental se puede describir de un modo más formal. Para ello, consideremos un conjunto $E = \{e_1, e_2, \ldots, e_{48}\}$ compuesto por los 48 estudiantes de la muestra; y una secuencia $P = \{p_1, p_2, \ldots, p_{10}\}$ de 10 problemas que tienen que resolver mediante HBPS. Cada vez que un estudiante $e_q \in E$ completa un problema $p_r \in P$, se almacena información de carácter afectivo sobre su estado relativo a la Valencia, la Activación y la Dominancia $\{V_{e_q,p_r}, A_{e_q,p_r}, D_{e_q,p_r}\}$, así como información sobre su rendimiento en la resolución del problema propuesto (S_{e_q}, p_r). La información afectiva se recogió a través de la autoevaluación del alumno mediante formularios SAM tras la finalización de cada problema. En la figura 7.2 se muestra el formulario utilizado para tal fin. El rendimiento del usuario se estimó a partir de la proporción de acciones correctas que no requirieron ningún tipo de ayuda durante la resolución del problema p_r propuesto, ya que una acción incorrecta suponía la provisión automática de una ayuda al estudiante.

La información recabada en la experimentación se empleó para la creación de dos clasificadores diferentes para la evaluación de cada una de las estrategias de maximización M_i descritas anteriormente, con el objetivo de entrenar las funciones

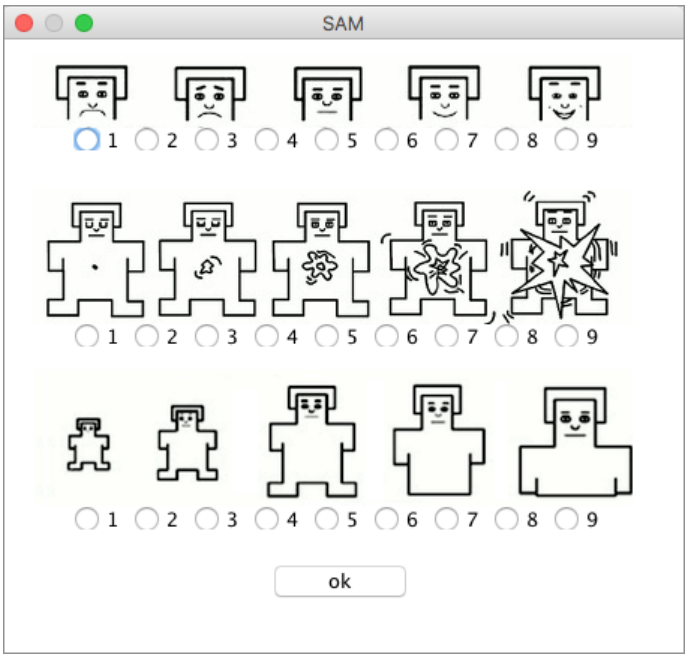


Figura 7.2: Test SAM utilizado para capturar los estados afectivos de los estudiantes

f'_{M_1} y f'_{M_2} que permitan al STI decidir el nivel de ayuda instruccional más adecuado a proporcionar al estudiante en función de la estrategia de maximización M_i establecida. Para ello, la información recolectada se usó para la construcción de un conjunto de muestras de entrenamiento con la forma $\{a, c_1, c_2, c_3\} \rightarrow L_M$, para cada posible combinación de estudiante y problema.

La ayuda a representa el nivel de concreción de la asistencia provista en un rango de valores enteros entre 1 y 4, correspondiendo el valor 1 al mínimo nivel de detalle a proporcionar al estudiante y el valor 4 al que mayor nivel de descripción ofrece. Las variables c_1, c_2 y c_3 representan la información de contexto de la situación instruccional que puede influir en la selección del nivel óptimo de ayuda a proveer. Estas variables contextuales se relacionan con la dificultad del problema que el alumno está resolviendo en un momento dado y con su nivel actual de competencia en la resolución de problemas aritméticos. En la tabla 7.1 se definen estas tres variables.

La dificultad del problema actual, c_1, se calcula como el rendimiento medio de todos los participantes, estimando el rendimiento individual de cada estudiante como la proporción de pasos de un problema resueltos sin ningún tipo de ayuda instruccional. Por otro lado, con el objetivo de paliar posibles diferencias entre la dificultad de los problemas, la competencia del estudiante se representa como una combinación de las variables contextuales c_2 y c_3. La primera de ellas, c_2, representa su rendimiento en la resolución del problema anterior. La segunda, c_3, la dificultad media del problema anterior, basada ésta en el rendimiento medio de

Dificultad del problema	
Dificultad del problema actual p_r	$c_1 = \frac{\sum_{q=1}^{N} S_{e_q,p_r}}{N}$

Competencia del usuario resolviendo problemas aritméticos	
Rendimiento del usuario en el problema anterior p_{r-1}	$c_2 = S_{e_q,p_{r-1}}$
Dificultad del problema anterior p_{r-1}	$c_3 = \frac{\sum_{q=1}^{N} S_{e_q,p_{r-1}}}{N}$

Tabla 7.1: Variables que definen el contexto para un estudiante particular $e_q \in E$ en la resolución de un problema concreto $p_r \in P$

todos los estudiantes.

El nivel de consecución L_{M_i} de una estrategia de maximización M_i se aproximó mediante las reglas definidas en la tabla 7.2, donde D representa los valores de Dominancia, V de Valencia y A de Activación. De este modo resultó posible medir el efecto de la aplicación de una estrategia de maximización concreta a través de la comparación de las valoraciones reportadas por el usuario en el test SAM antes y después de resolver el problema.

Para tratar de minimizar las posibles diferencias de sensibilidad de los usuarios en la valoración de las variaciones percibidas en su estado afectivo, éstas fueron convertidas a un valor discreto dentro del conjunto $\{+1, 0, -1\}$, correspondiendo el valor $+1$ a una variación positiva de la dimensión emocional evaluada en el test SAM (Dominancia, Activación o Valencia), -1 a una variación negativa y 0 a una neutral o desconocida. No obstante, únicamente las muestras que causaron variaciones en el estado afectivo (valores $+1$ o -1) fueron consideradas para el conjunto de entrenamiento, excluyéndose aquellas que no mostraron variación alguna (valor 0). Esta selección de muestras hizo que la aproximación de la función del nivel de consecución L_{M_i} se redujera a un problema de aprendizaje binario, resuelto, en este caso, mediante máquinas de soporte vectorial (SVM) con kernel radial y decisión probabilística, de forma que la salida de la SVM obtiene, en lugar de una predicción basada en una única clase, una probabilidad de pertenencia a cada una de las dos posibles clases $\{+1, -1\}$.

Es importante destacar que tanto las fórmulas definidas en la tabla 7.1 como las reglas de la tabla 7.2 utilizan en sus cálculos valores y valoraciones relativas al problema anterior ($D_{e_q,p_{r-1}}, V_{e_q,p_{r-1}}$ y $A_{e_q,p_{r-1}}$). Esto supone que la información del primer problema no será considerado para la formación del conjunto de entrenamiento al no disponer de información previa. De modo similar, las muestras

Estrategia	**Nivel de consecución de la estrategia de maximización**
M_1	$L_{M_1} = \begin{cases} +1 & \text{if } D_{e_q,p_r} > D_{e_q,p_{r-1}} \\ -1 & \text{if } D_{e_q,p_r} < D_{e_q,p_{r-1}} \\ 0 & \text{if } D_{e_q,p_r} = D_{e_q,p_{r-1}} \end{cases}$
M_2	$L_{M_2} = \begin{cases} +1 & \text{Si } V_{e_q,p_r} > V_{e_q,p_{r-1}} \text{ y } A_{e_q,p_r} > A_{e_q,p_{r-1}} \\ -1 & \text{Si } V_{e_q,p_r} < V_{e_q,p_{r-1}} \text{ y } A_{e_q,p_r} < A_{e_q,p_{r-1}} \\ 0 & \text{en cualquier otro caso} \end{cases}$

Tabla 7.2: Nivel de consecución de las estrategias descritas para un estudiante particular $e_q \in E$ en la resolución de un problema concreto $p_r \in P$

en las que no se prestó ningún tipo de ayuda instruccional al alumno tampoco fueron consideradas debido a que el nivel de ayuda fijado para el problema resultó irrelevante en estos casos. Estas consideraciones hicieron que el conjunto total de muestras de entrenamiento disponibles se vieran reducidas a 234 para el caso de f'_{M_1} y 84 para f'_{M_2}.

Por otro lado, como se pudo apreciar en la tabla 7.1, la información relativa al estado afectivo del usuario no forma parte del contexto. Esto supone que, una vez entrenado el sistema de aprendizaje que aproxima las funciones f'_{M_1} y f'_{M_2}, en un entorno de producción se podrá prescindir de los test SAM de autoevaluación, debido a que las valoraciones afectivas del usuario únicamente se emplean en la estimación de las funciones L_{M_1} y L_{M_2} que se pretenden predecir. Una vez que las funciones f'_{M_1} y f'_{M_2} han sido aproximadas mediante aprendizaje, la estimación de las funciones L_{M_1} y L_{M_2} sólo necesitan la información de contexto $\{c_1, c_2, c_3\}$, sin dependencia alguna sobre las variables afectivas consideradas.

Con todo ello, se crearon dos clasificadores para la evaluación del nivel de consecución L_{M_i} para cada estrategia de maximización M_i. El clasificador para la estrategia de maximización M_1 (Dominancia) se entrenó con los datos de entrada definidos en la tabla 7.3, mientras que el clasificador para M_2 (Valencia y Activación) lo hizo con los datos de la tabla 7.4.

La información de entrada necesaria en la fase de predicción se muestra en la tabla 7.5. La salida de cada clasificador corresponde a la probabilidad de pertenencia a cada una de las dos posibles clases $\{+1, -1\}$ sobre la dimensión

afectiva para la que se diseñó (Dominancia o Valencia y Activación) y para cada posible nivel de ayuda a susceptible a proveer al alumno para el problema propuesto por el STI. De este modo, la valoración del impacto afectivo que el establecimiento de un determinado nivel de ayuda a puede suponer se realizará mediante la ejecución del clasificador tantas veces como niveles de ayuda existan, variando en cada ejecución el nivel de ayuda propuesto (variable *proposed_help_level*) y valorando las probabilidades obtenidas para cada nivel de ayuda.

Variable	Entrada	Descripción
a	*help_level*	Nivel de ayuda facilitado al usuario
c_2	*previous_skill*	Nivel de conocimiento adquirido por el usuario en el problema anterior, con valores en el intervalo [0,1]
	previous_p_help_requests_per_step	Porcentaje de ayudas solicitadas por paso en el problema anterior
	previous_p_steps_with_help	Porcentaje de pasos con solicitud de ayuda realizadas en el problema anterior
c_3	*avg_skill_previous_problem*	Nivel de conocimiento medio adquirido por todos los usuarios para el problema anterior
	avg_p_help_requests_per_step_previous_problem	Porcentaje medio de ayudas solicitadas en cada paso por todos los usuarios para el problema anterior
	avg_p_steps_with_help_previous_problem	Porcentaje medio de pasos con solicitud de ayuda realizadas por todos los usuarios para el problema anterior
c_1	*avg_skill_current_problem*	Nivel de conocimiento medio adquirido por todos los usuarios para el problema actual
	avg_p_help_requests_per_step_current_problem	Porcentaje medio de ayudas solicitadas en cada paso por todos los usuarios para el problema actual
	avg_p_steps_with_help_current_problem	Porcentaje medio de pasos con solicitud de ayuda realizadas por todos los usuarios para el problema actual
Etiqueta	*autonomy_change_binarized*	Cambio en la Dominancia reportada por el usuario tras la realización del problema, con valores {-1, +1}

Tabla 7.3: Entradas del clasificador para la estrategia M_1 (Dominancia) durante la fase de entrenamiento

Variable	**Entrada**	**Descripción**
a	*help_level*	Nivel de ayuda facilitado al usuario
c_2	*previous_skill*	Nivel de conocimiento adquirido por el usuario en el problema anterior, con valores en el intervalo [0,1]
	previous_p_help_requests_per_step	Porcentaje de ayudas solicitadas por paso en el problema anterior
	previous_p_steps_with_help	Porcentaje de pasos con solicitud de ayuda realizadas en el problema anterior
c_3	*avg_skill_previous_problem*	Nivel de conocimiento medio adquirido por todos los usuarios para el problema anterior
	avg_p_help_requests_per_step_previous_problem	Porcentaje medio de ayudas solicitadas en cada paso por todos los usuarios para el problema anterior
	avg_p_steps_with_help_previous_problem	Porcentaje medio de pasos con solicitud de ayuda realizadas por todos los usuarios para el problema anterior
c_1	*avg_skill_current_problem*	Nivel de conocimiento medio adquirido por todos los usuarios para el problema actual
	avg_p_help_requests_per_step_current_problem	Porcentaje medio de ayudas solicitadas en cada paso por todos los usuarios para el problema actual
	avg_p_steps_with_help_current_problem	Porcentaje medio de pasos con solicitud de ayuda realizadas por todos los usuarios para el problema actual
Etiqueta	*valence_and_activation_change_binarized*	Cambios en la Valencia y en la Activación reportados por el usuario tras la realización del problema, con valores {-1, +1}

Tabla 7.4: Entradas del clasificador para la estrategia M_2 (Valencia y Activación) durante la fase de entrenamiento

Variable	Entrada	Descripción
a	*proposed_help_level*	Nivel de ayuda evaluado por el STI
c_2	*previous_skill*	Nivel de conocimiento adquirido por el usuario en el problema anterior, con valores en el intervalo [0,1]
	previous_p_help_requests_per_step	Porcentaje de ayudas solicitadas por paso en el problema anterior
	previous_p_steps_with_help	Porcentaje de pasos con solicitud de ayuda realizadas en el problema anterior
c_3	*avg_skill_previous_problem*	Nivel de conocimiento medio adquirido por todos los usuarios para el problema anterior
	avg_p_help_requests_per_step_previous_problem	Porcentaje medio de ayudas solicitadas en cada paso por todos los usuarios para el problema anterior
	avg_p_steps_with_help_previous_problem	Porcentaje medio de pasos con solicitud de ayuda realizadas por todos los usuarios para el problema anterior
c_1	*avg_skill_current_problem*	Nivel de conocimiento medio adquirido por todos los usuarios para el problema actual
	avg_p_help_requests_per_step_current_problem	Porcentaje medio de ayudas solicitadas en cada paso por todos los usuarios para el problema actual
	avg_p_steps_with_help_current_problem	Porcentaje medio de pasos con solicitud de ayuda realizadas por todos los usuarios para el problema actual

Tabla 7.5: Entradas de los clasificadores durante la fase de predicción

7.4. Implementación y análisis de los clasificadores afectivos

Evaluados diferentes métodos de clasificación, entre ellos vecino más próximo, perceptrón multicapa, árboles de decisión y máquinas de soporte vectorial, con los resultados en términos de cobertura y área ROC bajo la curva mostrados en la tabla 7.6, se optó por una implementación basada en máquinas de soporte vectorial con kernel radial mediante la librería LibSVM para la construcción de los clasificadores afectivos descritos en el apartado anterior. Los parámetros para la SVM con kernel RBF, optimizados mediante una estrategia de búsqueda *Grid Search*, y la precisión obtenida mediante un entrenamiento de tipo *leave-one-out cross-validation* para cada uno de los clasificadores, se detallan en la tabla 7.7.

Método evaluado	**Cobertura media y área ROC**	
	Dominancia(f'_{M_1})	**Valencia y Activación(f'_{M_2})**
SVM	64,50 % (0,694)	67,90 % (0,746)
K-NN	63,70 % (0,625)	66,70 % (0,649)
LMT	64,10 % (0,668)	66,70 % (0,717)
C4.5	65,40 % (0,629)	65,50 % (0,465)
Perceptrón multicapa	65,00 % (0,677)	61,90 % (0,591)

Tabla 7.6: Rendimiento obtenido por los diferentes métodos de clasificación evaluados para la construcción de los clasificadores afectivos de HBPS, en términos de cobertura media y área ROC

Clasificador SVM	**C**	**G**	**Precisión**	**ROC**
DOMINANCIA (f'_{M_1})	10	0,01	64,53 %	0,694
VALENCIA y ACTIVACIÓN (f'_{M_2})	10	0,01	67,86 %	0,746

Tabla 7.7: Optimización de los clasificadores afectivos: Dominancia (f'_{M_1}), Valencia y Activación (f'_{M_2})

En las figuras 7.3 y 7.4 se muestran las áreas ROC obtenidas para los clasificadores implementados para la aproximación de las funciones f'_{M_1} y f'_{M_2}, con valores para el área bajo la curva de 0,694 y 0,746, respectivamente. Estos resultados mantienen la hipótesis de que el conjunto $\{a, c_1, c_2, c_3\}$ puede utilizarse como predictor para las variables L_{M_1} y L_{M_2} y, por tanto, para determinar el nivel de ayuda a más apropiado para una situación descrita por su información contextual $\{c_1, c_2, c_3\}$.

Las matrices de confusión obtenidas durante la evaluación de los clasificadores para f'_{M_1} y f'_{M_2} se muestran en las tablas 7.8 y 7.9, respectivamente.

Por último, las medidas-F, cuyos valores representan el rendimiento de los clasificadores implementados mediante la relación entre precisión y cobertura, alcanzaron valores para $f'_{M_1} = 0,65$ y $f'_{M_2} = 0,68$. Estos valores indican que las

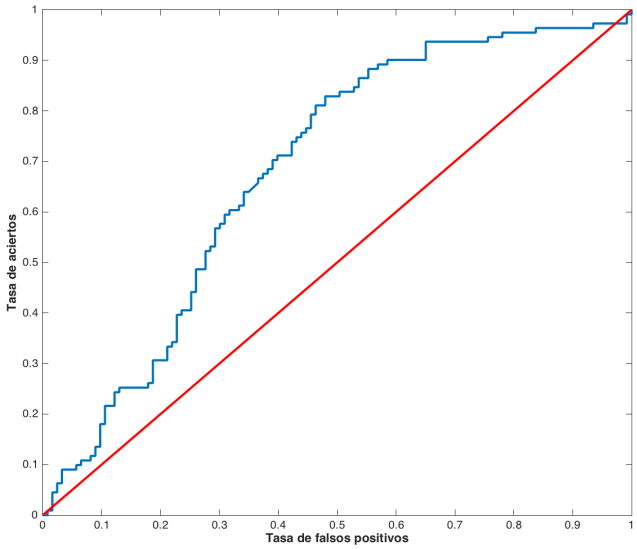


Figura 7.3: Área ROC (0,694) obtenida para el clasificador que aproxima f'_{M_1} (Dominancia)

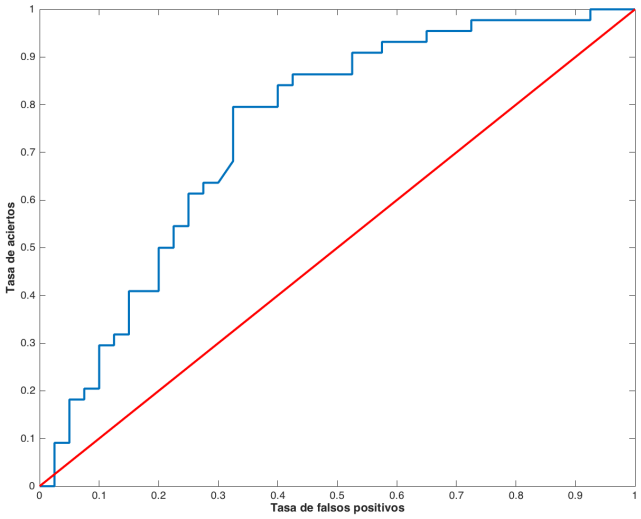


Figura 7.4: Área ROC (0,746) obtenida para el clasificador que aproxima f'_{M_2} (Valencia y Activación)

funciones aprendidas son capaces de predecir correctamente la variación de las observaciones en dos de cada tres intentos, en promedio.

	Predicciones		
	Clase -1	**Clase +1**	**Cobertura**
Clase -1 ($n = 111$)	73	38	**65,8 %**
Clase +1 ($n = 123$)	45	78	**63,4 %**
Verdaderos positivos	**65,8 %**	**63,4 %**	**67,9 %**
Falsos positivos	**34,2 %**	**36,6 %**	**32,2 %**

Tabla 7.8: Matriz de confusión obtenida en la aproximación de f'_{M_1} (Dominancia)

	Predicciones		
	Clase -1	**Clase +1**	**Cobertura**
Clase -1 ($n = 44$)	30	14	**68,2 %**
Clase +1 ($n = 40$)	13	27	**67,5 %**
Verdaderos positivos	**68,2 %**	**67,5 %**	**64,5 %**
Falsos positivos	**31,8 %**	**32,5 %**	**35,5 %**

Tabla 7.9: Matriz de confusión obtenida en la aproximación de f'_{M_2} (Valencia y Activación)

7.5. Experimentación

Los resultados expuestos en al apartado anterior proporcionan evidencias de que el nivel de ayuda proporcionado por el STI durante la resolución de un problema tiene consecuencias en el estado afectivo del estudiante, valorado éste sobre las dimensiones de la Dominancia, la Valencia y la Activación, y que la variación de estas variables puede ser predicha con una precisión razonable mediante el uso de información acerca del contexto del usuario y del problema que está resolviendo. En este sentido, con el propósito de poder validar la hipótesis de que el nivel de ayuda supone un efecto sobre el estado afectivo del estudiante, se llevó a cabo una nueva experimentación con 46 estudiantes de quinto curso de educación Primaria, con edades comprendidas entre los 10 y 11 años, que no habían participado en la experimentación previa diseñada para la recolección de datos.

Para llevar a cabo esta experimentación, se creó una nueva versión de HBPS para la resolución de los mismos 10 problemas aritméticos diseñados y empleados en la experimentación anterior. En esta nueva versión, el nivel de ayuda para el primer problema se estableció con el valor 1 (mínimo nivel de detalle de la ayuda). Los 9 problemas restantes fueron divididos en 3 conjuntos de 3 problemas cada uno. A cada uno de estos conjuntos se le aplicó una estrategia diferente en la decisión del nivel de ayuda a proporcionar por el STI: 1) en uno de ellos el nivel de ayuda se estableció de forma aleatoria ($M_{aleatoria}$); 2) a cada uno de los dos conjuntos restantes se le aplicó una de las dos estrategias de maximización M_i, de modo que el nivel de ayuda se decidió mediante el uso de las funciones entrenadas f'_{M_1} y f'_{M_2}. En todos los casos, el nivel de ayuda escogido se mantuvo constante para todo el problema.

Dado que en esta experimentación se emplearon los mismos 10 problemas

aritméticos utilizados para la creación del núcleo afectivo, resultó posible precalcular las dificultades asociadas para cada uno de ellos (parámetros c_1 y c_3). Por otro lado, el parámetro c_2, al estar basado exclusivamente en información relacionada con el rendimiento del estudiante en el último problema completado, se calculó en tiempo real.

Los test de autoevaluación SAM no fueron eliminados del STI para poder analizar el efecto de las decisiones tomadas por el núcleo afectivo sobre los niveles afectivos de los estudiantes.

Los 46 estudiantes fueron divididos en 6 grupos de similar tamaño. Sobre cada uno de estos grupos se aplicaron las tres estrategias descritas anteriormente (M_1, M_2 o $M_{aleatoria}$), pero en distinto orden, de modo que a cada grupo se le aplicó una estrategia diferente para la decisión de la ayuda instruccional a proporcionar durante la resolución de cada conjunto de 3 problemas. La distribución de los estudiantes en grupos y la secuencia de aplicación de las tres estrategias de maximización para la provisión de las ayudas instruccionales se describen en la tabla 7.10. A modo de ejemplo, al grupo número 2 se le aplicó la estrategia que persiguió maximizar la Dominancia (f'_{M_1}) durante la resolución de los problemas 2, 3 y 4; a los siguientes tres problemas se les aplicó una estrategia aleatoria ($f_{aleatoria}$) para la selección del nivel de ayuda a proveer; a los últimos tres problemas se les aplicó la estrategia para la maximización de la Valencia y la Activación conjunta (f'_{M_2}).

Id. del grupo	**Número de estudiantes**	**Problemas 2, 3 y 4**	**Problemas 5, 6 y 7**	**Problemas 8, 9 y 10**
1	7	f'_{M_1}	f'_{M_2}	$f_{aleatoria}$
2	8	f'_{M_1}	$f_{aleatoria}$	f'_{M_2}
3	8	f'_{M_2}	f'_{M_1}	$f_{aleatoria}$
4	8	f'_{M_2}	$f_{aleatoria}$	f'_{M_1}
5	8	$f_{aleatoria}$	f'_{M_1}	f'_{M_2}
6	7	$f_{aleatoria}$	f'_{M_2}	f'_{M_1}

Tabla 7.10: Aplicación de las estrategias de maximización para cada grupo de estudiantes y conjunto de problemas

Para llevar a cabo la experimentación fueron creadas *ad-hoc* seis imágenes *live* de GNU/Linux basadas en la distribución Ubuntu 14.04 LTS de 64 bits, con escritorio XFCE[1] y con soporte para la persistencia de datos. Las imágenes fueron grabadas en unidades de memoria USB 3.0 de 16 GB y se configuraron para la ejecución automática de la versión adaptada de HBPS que integraba los clasificadores afectivos implementados. Cada una de estas imágenes se configuró para la aplicación de las estrategias de maximización en la secuencia descrita en la tabla 7.10. Se reservó un 40 % de su capacidad total para el almacenamiento de los datos recogidos durante la experimentación.

[1] http://xubuntu.org

7.6. Análisis de resultados

Los datos recogidos durante la experimentación y almacenados en las memorias USB, se recopilaron y analizaron con el propósito de comparar el efecto de la aplicación de cada estrategia M_i sobre las dimensiones afectivas consideradas: Dominancia; Valencia y Activación.

En la tabla 7.11 se muestra el impacto de la aplicación de las funciones aprendidas f'_{M_1} y f'_{M_2} y la aplicación de una estrategia aleatoria $f_{aleatoria}$, sobre los valores de Dominancia reportados por los estudiantes en los test SAM. En este caso, la función f'_{M_1} para la maximización de la Dominancia (M_1) se aplicó en 116 ocasiones con la intención de incrementar el nivel de autonomía del estudiante. Los resultados mostraron que en un 43,10 % se consiguió un aumento en los niveles de autonomía, mientras que en un 31,04 % no hubo ningún cambio significativo y en un 25,86 % se obtuvo un resultado contrario a lo esperado. Esto contrasta con los resultados obtenidos en la aplicación de las funciones f'_{M_2} y $f_{aleatoria}$. En estos casos el número de ocasiones en las que se empleó fue similar, pero el efecto sobre la autonomía reportada por los estudiantes estuvo más balanceado, sin que se apreciara un efecto en la aplicación de estas estrategias cuando la intención es actuar sobre la Dominancia.

Estrategia aplicada	$\boldsymbol{\Delta D > 0}$ $\boldsymbol{(L_{M_1} = +1)}$	$\boldsymbol{\Delta D = 0}$ $\boldsymbol{(L_{M_1} = 0)}$	$\boldsymbol{\Delta D < 0}$ $\boldsymbol{(L_{M_1} = -1)}$	**Número de muestras**
f'_{M_1}	$\mathbf{43,10}\,\%$	$31,04\,\%$	$25,86\,\%$	116
f'_{M_2}	$30,97\,\%$	$32,75\,\%$	$36,28\,\%$	113
$f_{aleatoria}$	$31,36\,\%$	$32,20\,\%$	$36,44\,\%$	118

Tabla 7.11: Variaciones reportadas en la Dominancia ($\Delta D = D_{e_q,p_r} - D_{e_q,p_{r-1}}$) para cada una de las estrategias empleadas durante la resolución de los problemas

Por otro lado, los resultados referentes al efecto de la aplicación de la estrategia para la maximización conjunta de la Valencia y la Activación (M_2) se muestran en la tabla 7.12. En ella se observa que la función f'_{M_2} que persigue la maximización de ambas variables afectivas tiene un rendimiento superior en comparación con el resto de funciones, además de presentar el doble de casos positivos que negativos. Las otras dos estrategias, sin embargo, presentaron más casos negativos que positivos.

Estrategia aplicada	$\boldsymbol{\Delta(V,A) > 0}$ $\boldsymbol{(L_{M_2} = +1)}$	**Otros** $\boldsymbol{(L_{M_2} = 0)}$	$\boldsymbol{\Delta(V,A) < 0}$ $\boldsymbol{(L_{M_2} = -1)}$	**Número de muestras**
f'_{M_1}	$12,07\,\%$	$71,55\,\%$	$16,38\,\%$	116
f'_{M_2}	$\mathbf{19,47}\,\%$	$71,68\,\%$	$8,85\,\%$	113
$f_{aleatoria}$	$9.32\,\%$	$72,88\,\%$	$17,80\,\%$	118

Tabla 7.12: Variaciones reportadas en la Valencia ($\Delta V = V_{e_q,p_r} - V_{e_q,p_{r-1}}$) y en la Activación ($\Delta A = A_{e_q,p_r} - A_{e_q,p_{r-1}}$) para cada una de las estrategias empleadas durante la resolución de los problemas

La información de las tablas 7.11 y 7.12 también revelan que el uso de una función f_{M_i} no supone un impacto en los resultados de la estrategia alternativa en comparación con el uso de una estrategia simplemente aleatoria. En particular, los resultados para f'_{M_2} y $f_{aleatoria}$ son muy similares respecto a la variación de la Dominancia, así como f'_{M_1} y $f_{aleatoria}$ presentan resultados muy próximos con respecto a las variaciones de Valencia y Activación simultáneas.

Con el propósito valorar la significación estadística de los resultados obtenidos y evaluar su generalidad, para cada estudiante se calculó la variación media del nivel de cumplimento de las estrategias M_1 y M_2 bajo tres condiciones diferentes: f_{M_1}, f_{M_2} y $f_{aleatoria}$.

Respecto al nivel de consecución de L_{M_1} (variaciones de Dominancia), un análisis ANOVA de medidas repetidas determinó que las diferencias entre las tres condiciones eran estadísticamente significativas ($F(2, 88) = 3,17; p < 0,05$). El test de Shapiro Wilk (normalidad) y la prueba de Levene (homocedasticidad) determinaron que las condiciones para la aplicación del análisis ANOVA se cumplían. Adicionalmente, se verificó mediante la prueba de Mauchly que la asunción de esfericidad no se violaba ($\chi^2(2) = 0,90; p = 0,1029$).

Seguidamente se realizó un análisis post hoc para comparar las relaciones entre las condiciones f_{M_1}, f_{M_2} y $f_{aleatoria}$, encontrándose una diferencia estadísticamente significativa en L_{M_1} entre las condiciones $f_{M_1}(M = 0,18; \sigma = 0,05)$ y $f_{aleatoria}(M = -0,02; \sigma = 0,08)$, con valores para la prueba $t(44) = -2,48; p < 0,05$ y tamaño del efecto $r = 0,35$. De modo similar, también se encontraron diferencias estadísticamente significativas para L_{M_1} entre las condiciones $f_{M_1}(M = 0,18; \sigma = 0,05)$ y $f_{M_2}(M = -0,02; \sigma = 0,06)$, con valores para la prueba $t(44) = -2,28; p < 0,05$ y tamaño del efecto $r = 0,32$. Por último, no se encontraron diferencias estadísticamente significativas para L_{M_1} entre las condiciones f_{M_2} y $f_{aleatoria}$, con valores para la prueba $t(44) = -0,04; p < 0,972$ y tamaño del efecto $r = 0,005$.

Respecto al nivel de consecución de L_{M_2} (variaciones simultáneas de Valencia y Activación), otro análisis ANOVA de medidas repetidas determinó que las diferencias entre las tres condiciones eran estadísticamente significativas ($F(2, 88) = 5,33; p < 0,01$). De nuevo, el test de Shapiro Wilk y la prueba de Levene determinaron que las condiciones para la aplicación de ANOVA se cumplían, así como la prueba de Mauchly corroboró que la asunción de esfericidad no se violaba ($\chi^2(2) = 0,90; p = 0,1054$).

De modo equivalente al análisis realizado para la Dominancia, se realizó un análisis post hoc para comparar las relaciones entre las tres condiciones con respecto a la Valencia y la Activación. Un primer contraste determinó que existían diferencias estadísticamente significativas para L_{M_2} entre las condiciones $f_{M_2}(M = 0,09; \sigma = 0,04)$ y $f_{aleatoria}(M = -0,07; \sigma = 0,04)$, con valores para la prueba $t(44) = -2,93; p < 0,01$ y tamaño del efecto $r = 0,40$. También se encontraron diferencias estadísticamente significativas en L_{M_2} entre las condiciones $f_{M_2}(M = 0,09; \sigma = 0,04)$ y $f_{M_1}(M = -0,09; \sigma = 0,05)$, con valores para la prueba $t(44) = -2,52; p < 0,05$ y tamaño del efecto $r = 0,36$. Por último, no se encontraron diferencias estadísticamente significativas para L_{M_2} entre las condiciones f_{M_1} y $f_{aleatoria}$, con valores para la prueba $t(44) = 0,14; p < 0,888$ y

tamaño del efecto $r = 0,02$.

7.7. Conclusiones y Discusión

En esta última aportación se ha presentado una aproximación de tipo *sensor-free*, basada en el análisis de patrones de comportamiento de los estudiantes durante su interacción con HBPS y en las valoraciones sobre su estado emocional en las dimensiones de Dominancia, Valencia y Activación, con el propósito de determinar el nivel de ayuda instruccional más adecuado a proveer durante la resolución de los problemas propuestos por el STI, para mejorar las dimensiones afectivas del estudiante mediante dos estrategias de maximización previamente establecidas, dotando, de este modo, a HBPS de capacidades de decisión de carácter afectivo.

Los resultados analizados sugieren que es posible intervenir en el estado afectivo del usuario mediante la aplicación de estrategias específicas que permitan ajustar la respuesta del STI, persiguiendo la mejora de sus variables afectivas.

Aunque el núcleo afectivo incorporado a HBPS se diseñó con el objetivo de maximizar la Dominancia y la Valencia y la Activación conjuntamente, mediante la definición de dos estrategias de maximización específicas, la aproximación presentada no se limita exclusivamente a la aplicación de estas dos intenciones, pudiéndose adaptar a otros propósitos siempre y cuando el nivel de consecución perseguido pueda ser modelado mediante un valor numérico o una etiqueta.

Una limitación del trabajo presentado es la asunción de que la intención o estrategia de maximización a perseguir se encuentra explícita. Sin embargo, la elección y definición de estas estrategias en escenarios educativos reales es una tarea no exenta de complejidad. Incluso en entornos presenciales, en los que el profesor o tutor también toma en consideración las variables afectivas de sus alumnos con el propósito de seleccionar acciones y explicaciones que proporcionen un balance entre el aprendizaje y su impacto emocional. Además, estas estrategias pueden variar en el tiempo, dependiendo de la situación y el contexto específico de aprendizaje. Aún así, el método propuesto proporciona la flexibilidad para modificar su comportamiento mediante la adaptación de las funciones de aprendizaje f'_{M_i}.

Otra limitación de la aproximación propuesta se relaciona con el hecho de que cada ayuda instruccional proporcionada al estudiante ha sido considerada de forma aislada. Posiblemente se podrían alcanzar mejoras superiores mediante la consideración de secuencias o combinaciones de ayudas instruccionales, aspecto que podría ser objeto de investigaciones futuras.

Por último, dos aspectos que podrían afectar al rendimiento de la aproximación propuesta es la selección del método de clasificación o regresión a emplear, así como las variables a considerar como parte del contexto. Estos aspectos se encuentran presentes en cualquier sistema de reconocimiento de patrones, pudiéndose aplicar las mismas reglas y métodos descritos en [149]. Cuanto mayor sea el número de variables a considerar más información de la situación se dispondrá a costa de incrementar la dimensionalidad del problema de clasificación a resolver. Determinar

las variables más adecuadas puede ser considerado como un problema de selección de características que debería ser estudiado en cada caso particular.

Parte IV

Conclusiones y trabajos futuros

Capítulo 8

Conclusiones y trabajos futuros

Resumen
En este último capítulo se resume el trabajo de investigación llevado a cabo y se exponen las conclusiones finales alcanzadas tras la realización del estudio presentado. Asimismo, considerando que en cualquier investigación realizada siempre quedan horizontes por explorar, se establecen las posibles líneas de investigación futuras que podrían dar continuidad a este trabajo de tesis.

Contenidos

8.1. Visión por computador 126
8.2. Señales fisiológicas 127
8.3. Patrones de comportamiento 128
8.4. Detección emocional en entornos de *e-learning* 129
8.5. Líneas de investigación futura 130
8.6. Publicaciones resultantes 131

La detección y evaluación del estado afectivo de un usuario constituye una tendencia de interés creciente en diversidad de entornos como el publicitario, el de la salud, los videojuegos o la educación, entre otros. No obstante, su detección, análisis y respuesta es una tarea compleja, no exenta de imprecisiones y limitada cuando se compara con la percepción realizada por un ser humano.

En este trabajo de tesis se ha realizado una revisión de la literatura más relevante sobre la detección y el análisis automático del estado afectivo de un usuario desde dos perspectivas diferenciadas: la detección emocional y sus aplicaciones en el ámbito educativo mediante sistemas adaptativos basados en la tutorización

inteligente. En esta revisión se han analizado las técnicas habitualmente empleadas en el reconocimiento emocional del usuario: visión por computador; análisis de señales fisiológicas; y reconocimiento de patrones de comportamiento asociados y sus relaciones con posibles estados afectivos subyacentes.

Por otro lado, para cada una de estas perspectivas, se han desarrollado diferentes aportaciones en la detección emocional y en la mejora instruccional.

8.1. Visión por computador

Los métodos de detección emocional basados en técnicas de visión por computador han sido tradicionalmente usados en sistemas de reconocimiento afectivo básicamente por dos razones: por imitación humana y por su baja intrusividad. Sus técnicas suelen basarse en modelos de representación facial 2D o 3D. Entre sus ventajas se encuentra la relativa facilidad para la validación de los métodos empleados. Por el contrario, entre sus inconvenientes se encuentra la susceptibilidad a las variaciones lumínicas, rotaciones de cabeza, oclusiones o, incluso, a diferencias entre sujetos tales como cambios en la textura de la piel, rasgos étnicos o de la edad.

Un requerimiento para el desarrollo de sistemas que sean capaces de detectar el estado afectivo del sujeto, es la disponibilidad de un corpus que pueda ser usado para entrenar y evaluar el rendimiento de los sistemas basados en el análisis de datos sobre emociones y estados afectivos. En este sentido se exploró la base de datos multimodal de expresiones faciales y emociones FEEDB. Este corpus contiene grabaciones de sujetos interpretando diferentes expresiones faciales. Las grabaciones fueron recogidas con un sensor Microsoft Kinect y es ofrecida a la comunidad investigadora como un repositorio abierto. Cada grabación está compuesta de canales independientes y sincronizados de color y profundidad y se proporcionan en el formato propietario entregado por Kinect: ficheros binarios XED. Aunque la base de datos incorpora un conjunto extenso de grabaciones y de emociones, su formato supone un *handicap* para cualquier desarrollador de sistemas afectivos que desee utilizarla para entrenar y evaluar una solución. Esto es así porque aunque los archivos XED pueden ser fácilmente procesados mediante diversas utilidades proporcionadas por Microsoft, es necesario para su uso disponer de un dispositivo hardware Microsoft Kinect como licencia de uso. Por este motivo, con el propósito de evitar la necesidad de disponer de un sensor Kinect, se desarrolló un sistema capaz de procesar las grabaciones en formato XED y de extraer la información relevante para la clasificación de emociones en formato texto, extendiendo, de este modo, las posibilidades ofrecidas en FEEDB.

La principal aportación realizada en este estudio de tesis sobre FEEDB ha sido la extracción de características faciales a partir de las grabaciones originales en el formato propietario XED de Microsoft, en concreto 100 características faciales, la posición de la cabeza y sus ángulos de inclinación, junto con 6 AU y 11 SU, según el modelo Candide-3 en el que se basa, y su almacenamiento en ficheros de texto independientes para una combinación total de 88 sujetos-emociones, con el propósito de que pueda ser utilizada por cualquier investigador sin la complejidad

de conocer la estructura de datos del formato XED y la necesidad de disponer de un sensor Kinect, con independencia de la plataforma hardware o software utilizada para el desarrollo de un sistema de detección afectiva y de forma sencilla al tener que procesar únicamente ficheros de texto. Esta aportación en modo de extensión abre nuevas posibilidades a la aplicación de toda la amplia variedad de técnicas de clasificación disponibles sin modificaciones (o con sencillas adaptaciones), lo que significa que se podrán concentrar los esfuerzos sobre los datos y sobre la extracción de conocimiento en lugar de buscar o adaptar algoritmos para que trabajen con estrictos formatos propietarios y, en ocasiones, con limitada accesibilidad.

Siguiendo la línea de investigación en la detección emocional mediante sistemas de visión artificial, en esta tesis se ha propuesto y desarrollado un método denominado *Eigenexpressions* como una solución holística basada en la apariencia para el reconocimiento de expresiones faciales, fundamentado en el método estándar de *Eigenfaces* y evaluado sobre la base de datos Cohn-Kanade+ para la detección de las seis emociones primarias, así como una extensión al propio método de Eigenexpressions basada en la creación de máscaras de expresiones faciales para la clasificación de expresiones.

Aunque con *Eigenexpressions* y con la extensión basada en máscaras se obtuvieron mejoras substanciales con respecto al método estándar *Eigenfaces* para el reconocimiento de expresiones faciales, al estar éstos basados en métodos de reducción basados en PCA es de prever que su rendimiento se vea mermado sobre un conjunto de rostros diferentes al conjunto de entrenamiento y evaluación utilizado y, de modo equivalente, en un entorno de interacción real donde las condiciones lumínicas pueden presentar grandes variaciones, así como rotaciones de cabeza que podrían afectar drásticamente a su rendimiento. No obstante, un elevado número de muestras disponibles en la fase de entrenamiento con diferentes condiciones lumínicas y posiciones de cabeza, unido al empleo de técnicas para reducir las variaciones de luz o corregir la posición de la cabeza, podría suponer que el sistema respondiera con mayor precisión incluso en entornos diferentes a los típicamente controlados en laboratorio. De modo similar, es factible considerar que la precisión de *Eigenexpressions* podría ser mejorada si para el análisis se hubiera podido disponer de varias fuentes de información simultánea. En este sentido, la solución desarrollada en este trabajo de tesis podría constituir un buen candidato para ser incorporado en aproximaciones multimodales, donde podría aportar información sobre la expresión facial detectada, pudiéndose utilizar como fuente de información complementaria junto a métodos basados en otros tipos de características.

8.2. Señales fisiológicas

Los métodos basados en el análisis de señales fisiológicas se encargan de analizar variables biométricas como el ritmo cardíaco, temperatura corporal, conductividad de la piel, EEG, etc. Su principal ventaja se encuentra en que permiten obtener información directa del sujeto difícilmente falsificable por el mismo. Adicionalmente, la combinación de varias de ellas puede aportar valiosa

información sobre el estado afectivo del sujeto como, por ejemplo, la evaluación del ritmo cardíaco junto con mediciones de la conductividad de la piel. Entre sus inconvenientes se encuentra que son técnicas más intrusivas que las basadas en visión –aspecto minimizado por la reciente aparición de prendas y objetos de uso cotidiano (*wereables*) capaces de monitorizar algunos parámetros de la actividad fisiológica de la persona que los viste–, requieren un instrumental más complejo y preciso, un tratamiento computacional más costoso, además de que pueden presentar una gran variación entre individuos, aspecto que las hacen complejas de analizar y clasificar mediante patrones que definan el estado afectivo del sujeto.

En lo que se refiere al estudio de este tipo de señales, la información proveniente de señales EEG incluida en la base de datos MAHNOB-HCI fue analizada con el objetivo de encontrar posibles patrones que pudieran identificar emociones subyacentes. Diferentes técnicas de clasificación sobre el conjunto de emociones recogidas en la base de datos, así como sobre su agrupación bajo las dimensiones de Activación y Valencia, fueron aplicadas a los datos EEG sin conseguir una precisión superior a la obtenida por los autores del corpus.

La conclusión más relevante a la que se llegó tras la experimentación llevada a cabo sobre las señales EEG contenidas en MAHNOB-HCI es que, teniendo en cuenta que la información recabada en la base de datos fue recogida en un entorno controlado de laboratorio, el tratamiento de este tipo de señales resulta complejo al presentar una alta variabilidad entre sujetos, aspecto que dificulta el reconocimiento del estado emocional del usuario. Este problema puede verse agravado en un entorno real, donde es de prever que la información capturada sea de peor calidad que la obtenida en un entorno controlado. Probablemente señales con menor complejidad y variabilidad entre sujetos, como pueden ser las recogidas mediante electrocardiogramas o las relativas a los patrones de respiración del usuario, entre otras, puedan ser combinadas junto con otras fuentes de información y aportar información relevante sobre estados emocionales como estrés, ansiedad o aburrimiento.

8.3. Patrones de comportamiento

Es evidente que existe una relación entre el comportamiento de un usuario cuando interactúa con un sistema y su estado afectivo y mental subyacente. En este sentido los patrones de comportamiento pueden dar respuesta al modo de interacción de un usuario con el sistema y a su estado afectivo, aprendiendo de sus hábitos y preferencias para poder prestarle proactivamente ayudas o servicios personalizados. Su principal problema es que sus aplicaciones son altamente dependientes del dominio y, por tanto, difícilmente extrapolables a otros contextos.

Desde el punto de vista de las técnicas comúnmente utilizadas para analizar posibles patrones de comportamiento, este método no se puede considerar por sí mismo como intrínsecamente autónomo, sino que se sirve de otras técnicas para poder recolectar información para analizar *a posteriori* la existencia de patrones de comportamiento asociados a estados afectivos concretos. Entre estas técnicas se encuentran las basadas en visión artificial, análisis de la voz, movimientos

corporales, señales fisiológicas, registros de interacción usuario-software como movimientos, velocidad y pulsaciones del ratón, del teclado o de ambos, registros de tiempos de respuesta, tiempo consumido en la resolución de una tarea, tasas de errores y aciertos cometidos, número y tipo de ayudas solicitadas, así como cuestionarios y auto-evaluaciones para reportar las percepciones y sentimientos durante su interacción con el sistema, como por ejemplo mediante formularios SAM que recojan la percepción del usuario sobre la Valencia, la Activación y la Dominancia.

Las técnicas de aprendizaje automático son ampliamente utilizadas en el análisis de patrones. En esta tesis se implementó un conjunto de clasificadores basados en máquinas de soporte vectorial para predecir la probabilidad asociada al estado afectivo del usuario sobre las dimensiones de Dominancia, Activación y Valencia. Estos clasificadores fueron incorporados posteriormente a un STI para el aprendizaje de la aritmética y el álgebra lineal, con el objetivo de poder regular el nivel de ayuda instruccional a proporcionar al alumno para mejorar su estado afectivo. Los resultados obtenidos en la experimentación llevada a cabo demostraron que es posible modificar las variables afectivas mediante el ajuste de la respuesta del STI. En este sentido, es factible aseverar que aunque el análisis de patrones de comportamiento es altamente dependiente del contexto, puede considerarse como una valiosa herramienta de ayuda para la predicción del estado afectivo del usuario que interactúa con el sistema.

8.4. Detección emocional en entornos de *e-learning*

Tras el estudio realizado en este trabajo de tesis acerca de los diferentes métodos y técnicas utilizados habitualmente en el análisis y reconocimiento de expresiones faciales, emociones y estados afectivos, se procedió a analizar el impacto que las emociones pueden suponer en el aprendizaje desde el punto de vista de los sistemas adaptativos basados en la tutorización inteligente. De este modo, un STI que incorpore soporte afectivo podría inferir en cada momento el estado afectivo del estudiante y adaptar su funcionamiento al mismo, con el propósito de aplicar un conjunto de estrategias instruccionales que minimicen posibles situaciones de aburrimiento, frustración o abandono y, por tanto, mejorar su rendimiento.

Para tal fin se diseñó un estudio experimental mediante el uso de un STI para el aprendizaje de la aritmética y el álgebra lineal denominado HBPS, basado en la provisión de diferentes tipos y niveles de ayuda como estrategia de soporte instruccional al alumno.

Una primera experimentación en un entorno educativo real sirvió para la recolección de información de comportamiento, así como de las valoraciones afectivas reportadas por los propios estudiantes mediante formularios SAM. Con esta información se entrenó un sistema de aprendizaje automático con el fin de prestar colaboración a HBPS en las decisiones del nivel de ayuda instruccional más adecuado a proveer al estudiante, tomando en consideración su modelo de conocimiento, el contexto formativo y su estado afectivo, con la intención de

que el STI pudiera intervenir en el estado afectivo del alumno a través de las ayudas provistas al mismo, sobre dos posibles dimensiones afectivas: por un lado la Dominancia; por otro la Valencia y la Activación conjuntamente.

El sistema de aprendizaje implementado fue incorporado a HBPS para dotar al STI de la capacidad para determinar en cada momento el nivel de ayuda instruccional más apropiado a facilitar al alumno en función de la dimensión afectiva que se deseara maximizar. La versión de HBPS con soporte afectivo fue evaluado en un entorno real mediante la aplicación de diferentes estrategias de intervención para cada conjunto de los problemas propuestos. Estas estrategias demostraron resultar efectivas desde el punto de vista de la regulación del estado afectivo del alumno.

La conclusión más relevante que se ha podido obtener en este trabajo, es que es posible intervenir en el estado afectivo del estudiante mediante la aplicación de diferentes estrategias instruccionales dirigidas a maximizar sus dimensiones afectivas. Es decir, analizando su modelo cognitivo, el contexto formativo y su estado afectivo, es posible proporcionar ayudas individualizadas que permitan al alumno no sólo alcanzar la solución al problema planteado, sino también intervenir en su estado afectivo, aspecto que podría influir en su aprendizaje final. Esto puede tener un potencial impacto en escenarios diferentes al educativo, como por ejemplo en el ámbito del *marketing*, donde se podría intervenir en el estado emocional de un usuario consumidor mediante la aplicación de diferentes estrategias orientadas a la modificación de su estado afectivo y, de este modo, condicionar sus preferencias hacia una línea determinada de bienes o servicios.

8.5. Líneas de investigación futura

El soporte afectivo incorporado a HBPS se ha centrado en la maximización de las dimensiones afectivas relativas a la Dominancia, la Valencia y la Activación, evaluando los cambios positivos o negativos producidos sobre estas variables. Desde esta perspectiva, podría resultar interesante aumentar la granularidad de las variaciones del estado afectivo de los usuarios, de modo que no sólo puedan valorarse y aplicarse estrategias de maximización sobre la dirección de los cambios en las variables afectivas, sino que se pudiera valorar cuantitativamente cada una de ellas con el objetivo de aportar un mayor nivel de detalle en las decisiones consideradas por el STI.

De modo similar, la identificación de patrones de comportamiento asociados a estados afectivos concretos y particulares en escenarios educativos, como el aburrimiento, la distracción o la frustración, podrían resultar de utilidad en la aplicación de estrategias de ayuda instruccional específicas en entornos de *e-learning*. De este modo, la intervención podría ser regulada combinando la identificación de estos otros estados afectivos con su correspondiente dimensión afectiva en términos de Dominancia, Valencia y Activación.

8.6. Publicaciones resultantes

Gran parte de los resultados de la investigación llevada a cabo descrita en esta tesis doctoral ha dado lugar a publicaciones en revistas y congresos internacionales en áreas relacionadas con la computación afectiva, el reconocimiento de patrones y la educación.

[120] Luis Marco-Giménez, Miguel Arevalillo-Herráez, and Cristina Cuhna-Pérez. Eigenexpressions: Emotion recognition using multiple eigenspaces. In *Ib-PRIA*, pages 758–765, 2013

[119] Luis Marco-Giménez, Miguel Arevalillo-Herráez, Aladdin Ayesh, and Mariusz Szwoch. An extension to the feedb multimodal database of facial expressions and emotions. *Eurosis ESM 2015*, page 455 – 460, 2015

[118] Luis Marco-Giménez, Miguel Arevalillo-Herráez, Francesc J. Ferri, Salvador Moreno-Picot, Jesus Boticario, Olga C. Santos, Sergio Salmeron-Majadas, Mar Saneiro, Raul Uria-Rivas, David Arnau, José Antonio González-Calero, Aladdin Ayesh, Raúl Cabestrero, Pilar Quirós, Pablo Arnau-González, and Naeem Ramzan. Affective and behavioral assessment for adaptive intelligent tutoring systems. In *Late-breaking Results, Posters, Demos, Doctoral Consortium and Workshops Proceedings of the 24th ACM Conference on User Modeling, Adaptation and Personalisation (UMAP 2016), Halifax, Canada, July 13-16, 2016.*, 2016

[7] Miguel Arevalillo-Herráez, David Arnau, and Luis Marco-Giménez. Domain-specific knowledge representation and inference engine for an intelligent tutoring system. *Knowledge-Based Systems*, 49:97–105, 2013

[8] Miguel Arevalillo-Herráez, David Arnau, Luis Marco-Giménez, José Antonio González-Calero, Salvador Moreno-Picot, Paloma Moreno-Clari, Aladdin Ayesh, Olga C. Santos, Jesus Boticario, Mar Saneiro, Sergio Salmeron-Majadas, Raúl Cabestrero, and Pilar Quirós. Providing personalized guidance in arithmetic problem solving. In *UMAP Workshops*, 2014

Bibliografía

[1] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, and N. Sebe. Decaf: Meg-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing*, 6(3):209–222, July 2015.

[2] J. Ahlberg. Candide-3. an updated parameterized face. Technical Report Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.

[3] M. Ainley. Connecting with learning: Motivation, affect and cognition in interest processes. *Educational Psychology Review 18*, pages 391—-405, 2006.

[4] P. Aishwarya and K. Marcus. Face recognition using multiple eigenface subspaces. *Journal of Engineering and Technology Research*, pages (28):139–143, 2010.

[5] Omar AlZoubi, Rafael A. Calvo, and Ronald H. Stevens. *AI 2009: Advances in Artificial Intelligence: 22nd Australasian Joint Conference, Melbourne, Australia, December 1-4, 2009. Proceedings*, chapter Classification of EEG for Affect Recognition: An Adaptive Approach, pages 52–61. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[6] Omar AlZoubi, Davide Fossati, Sidney D'Mello, and Rafael A Calvo. Affect detection from non-stationary physiological data using ensemble classifiers. *Evolving Systems*, 6(2):79–92, 2015.

[7] Miguel Arevalillo-Herráez, David Arnau, and Luis Marco-Giménez. Domain-specific knowledge representation and inference engine for an intelligent tutoring system. *Knowledge-Based Systems*, 49:97–105, 2013.

[8] Miguel Arevalillo-Herráez, David Arnau, Luis Marco-Giménez, José Antonio González-Calero, Salvador Moreno-Picot, Paloma Moreno-Clari, Aladdin Ayesh, Olga C. Santos, Jesus Boticario, Mar Saneiro, Sergio Salmeron-Majadas, Raúl Cabestrero, and Pilar Quirós. Providing personalized guidance in arithmetic problem solving. In *UMAP Workshops*, 2014.

[9] Roger Azevedo and Amber Chauncey Strain. *Integrating Cognitive, Metacognitive, and Affective Regulatory Processes with MetaTutor*, pages 141–154. Springer New York, New York, NY, 2011.

[10] Asier Aztiria, Alberto Izaguirre, and Juan Carlos Augusto. Learning patterns in ambient intelligence environments: a survey. *Artificial Intelligence Review*, 34(1):35–51, 2010.

[11] Areej Babiker, Ibrahima Faye, and Aamir Malik. Pupillary behavior in positive and negative emotions. In *Signal and Image Processing Applications (ICSIPA), 2013 IEEE International Conference on*, pages 379–383. IEEE, 2013.

[12] Jeremy N. Bailenson, Emmanuel D. Pontikakis, Iris B. Mauss, James J. Gross, Maria E. Jabon, Cendri A.C. Hutcherson, Clifford Nass, and Oliver John. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human-Computer Studies*, 66(5):303 – 317, 2008.

[13] Ryan Shaun Baker, Albert T. Corbett, Kenneth R. Koedinger, and Ido Roll. *Detecting When Students Game the System, Across Tutor Subjects and Classroom Cohorts*, pages 220–224. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[14] Tanja Bänziger, Marcello Mortillaro, and Klaus R Scherer. Introducing the geneva multimodal expression corpus for experimental research on emotion perception. *Emotion*, 12(5):1161, 2012.

[15] Marian Stewart Bartlett, Joseph C Hager, Paul Ekman, and Terrence J Sejnowski. Measuring facial expressions by computer image analysis. *Psychophysiology*, 36(02):253–263, 1999.

[16] Marian Stewart Bartlett, Gwen Littlewort, Mark Frank, Claudia Lainscsek, Ian Fasel, and Javier Movellan. Recognizing facial expression: machine learning and application to spontaneous behavior. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 568–573. IEEE, 2005.

[17] Marian Stewart Bartlett, Gwen Littlewort, Mark Frank, Claudia Lainscsek, Ian Fasel, and Javier Movellan. Fully automatic facial action recognition in spontaneous behavior. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 223–230. IEEE, 2006.

[18] Marian Stewart Bartlett, Gwen Littlewort, Claudia Lainscsek, Ian Fasel, and Javier Movellan. Machine learning methods for fully automatic recognition of facial expressions and facial actions. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 592–597. IEEE, 2004.

[19] MS Bartlett, GC Littlewort, TJ Sejnowski, and JR Movellan. A prototype for automatic recognition of spontaneous facial actions. In *Advances in Neural Information Processing Systems*, pages 1295–1302, 2003.

[20] A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. How to find trouble in communication. *Speech Communication*, 40:117–143, 2003.

[21] M. Parisa Beham and S. Mohamed Mansoor Roomi. Article: Face recognition using appearance based approach: A literature survey. *IJCA Proceedings on*

International Conference and workshop on Emerging Trends in Technology (ICWET 2012), icwet(12):16–21, March 2012. Full text available.

[22] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720, July 1997.

[23] Robert Bixler and Sidney D'Mello. Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 225–234. ACM, 2013.

[24] Koen B. E. Böcker, Jurgen A. G. van Avermaete, and Margaretha M. C. van den Berg-Lenssen. The international 10–20 system revisited: Cartesian and spherical co-ordinates. *Brain Topography*, 6(3):231–235, 1994.

[25] Danny Oude Bos. Eeg-based emotion recognition. *The Influence of Visual and Auditory Stimuli*, pages 1–17, 2006.

[26] Nigel Bosch, Yuxuan Chen, and Sidney D'Mello. It's written on your face: detecting affective states from facial expressions while learning computer programming. In *International Conference on Intelligent Tutoring Systems*, pages 39–44. Springer, 2014.

[27] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49 – 59, 1994.

[28] Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.

[29] Serge Brand, Torsten Reimer, and Klaus Opwis. How do we learn in a negative mood? effects of a negative mood on transfer and learning. *Learning and instruction*, 17(1):1–16, 2007.

[30] W. Burgin, C. Pantofaru, and W.D. Smart. Using depth information to improve face detection. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*, pages 119–120, March 2011.

[31] Kwang-Sub Byun, Chang-Hyun Park, and Kwee-Bo Sim. Emotion recognition from facial expression using hybrid-feature extraction. In *SICE 2004 Annual Conference*, volume 3, pages 2483–2487 vol. 3, Aug 2004.

[32] R.A. Calvo and S. D'Mello. Affect detection: An interdisciplinary review of models, methods, and their applications. *Affective Computing, IEEE Transactions on*, 1(1):18–37, Jan 2010.

[33] George Caridakis, Lori Malatesta, Loic Kessous, Noam Amir, Amaryllis Raouzaiou, and Kostas Karpouzis. Modeling naturalistic affective states via facial and vocal expressions recognition. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 146–154. ACM, 2006.

[34] Sandra Carvalho, Jorge Leite, Santiago Galdo-Álvarez, and Óscar F. Gonçalves. The emotional movie database (emdb): A self-report and psychophysiological study. *Applied Psychophysiology and Biofeedback*, 37(4):279–294, 2012.

[35] Ginevra Castellano, Santiago D. Villalba, and Antonio Camurri. *Affective Computing and Intelligent Interaction: Second International Conference, ACII 2007 Lisbon, Portugal, September 12-14, 2007 Proceedings*, chapter Recognising Human Emotions from Body Movement and Gesture Dynamics, pages 71–82. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[36] Ya Chang, Changbo Hu, Rogerio Feris, and Matthew Turk. Manifold based analysis of facial expression. *Image and Vision Computing*, 24(6):605–614, 2006.

[37] Ya Chang, Changbo Hu, and M. Turk. Probabilistic expression analysis on manifolds. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–520–II–527 Vol.2, June 2004.

[38] Ya Chang, Marcelo Vieira, Matthew Turk, and Luiz Velho. Automatic 3d facial expression analysis in videos. In *International Workshop on Analysis and Modeling of Faces and Gestures*, pages 293–307. Springer, 2005.

[39] Ira Cohen, Nicu Sebe, Ashutosh Garg, Lawrence S Chen, and Thomas S Huang. Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and image understanding*, 91(1):160–187, 2003.

[40] Jeffrey F Cohn and Karen L Schmidt. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(02):121–132, 2004.

[41] Cristina Conati and Heather Maclaren. Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3):267–303, 2009.

[42] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero. Survey on rgb, 3d, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):1548–1568, Aug 2016.

[43] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[44] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, January 1967.

[45] Ryan SJ d Baker, Sujith M Gowda, Michael Wixon, Jessica Kalka, Angela Z Wagner, Aatish Salvi, Vincent Aleven, Gail W Kusbit, Jaclyn Ocumpaugh, and Lisa Rossi. Towards sensor-free affect detection in cognitive tutor algebra. *International Educational Data Mining Society*, 2012.

[46] Dragoş Datcu and Léon Rothkrantz. Facial expression recognition in still pictures and videos using active appearance models: A comparison approach.

In *Proceedings of the 2007 International Conference on Computer Systems and Technologies*, CompSysTech '07, pages 112:1–112:6, New York, NY, USA, 2007. ACM.

[47] Richard J Davidson. Anterior cerebral asymmetry and the nature of emotion. *Brain and cognition*, 20(1):125–151, 1992.

[48] Richard J. Davidson. Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology*, 40(5):655–665, 2003.

[49] J. Ruiz del Solar and P. Navarrete. Eigenspace-based face recognition: a comparative study of different approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35(3):315–325, Aug 2005.

[50] Sidney D'Mello and Art Graesser. Automatic detection of learner's affect from gross body language. *Appl. Artif. Intell.*, 23(2):123–150, February 2009.

[51] Sidney K D'Mello. Emotional rollercoasters: Day differences in affect incidence during learning. In *FLAIRS Conference*, 2014.

[52] Sidney K. D'Mello, Scotty D. Craig, and Art C. Graesser. Multimethod assessment of affective experience and expression during deep learning. *Int. J. Learn. Technol.*, 4(3/4):165–187, October 2009.

[53] Sidney K D'Mello, Scotty D Craig, Amy Witherspoon, Bethany Mcdaniel, and Arthur Graesser. Automatic detection of learner's affect from conversational cues. *User modeling and user-adapted interaction*, 18(1-2):45–80, 2008.

[54] Sidney K. D'Mello and Arthur Graesser. Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, 20(2):147–187, 2010.

[55] Sidney K D'mello and Jacqueline Kory. A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR)*, 47(3):43, 2015.

[56] Gianluca Donato, Marian Stewart Bartlett, Joseph C. Hager, Paul Ekman, and Terrence J. Sejnowski. Classifying facial actions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(10):974–989, October 1999.

[57] Ellen Douglas-Cowie, Cate Cox, Jean-Claude Martin, Laurence Devillers, Roddy Cowie, Ian Sneddon, Margaret McRorie, Catherine Pelachaud, Christopher Peters, Orla Lowry, et al. The humaine database. In *Emotion-Oriented Systems*, pages 243–284. Springer, 2011.

[58] M.L.I.C. Dy, I.V.L. Espinosa, P.P.V. Go, C.M.M. Mendez, and J.W. Cu. Multimodal emotion recognition using a spontaneous filipino emotion database. In *Human-Centric Computing (HumanCom), 2010 3rd International Conference on*, pages 1–5, Aug 2010.

[59] Gareth J. Edwards, Timothy F. Cootes, and Christopher J. Taylor. Face recognition using active appearance models. In *Proceedings of the 5th*

European Conference on Computer Vision-Volume II - Volume II, ECCV '98, pages 581–595, London, UK, UK, 1998. Springer-Verlag.

[60] P. Ekman and W. Friesen. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, Palo Alto, 1978.

[61] P. Ekman, J. Rosenberg, and Hager J. Facial action coding system affect interpretation dictionary (facsaid), 1998.

[62] Paul Ekman. *Emotions revealed: recognizing faces and feelings to improve communication and emotional life.* Times Books, New York, 2003.

[63] Paul Ekman and Wallace V Friesen. Nonverbal leakage and clues to deception. *Psychiatry*, 32(1):88–106, 1969.

[64] Rana El Kaliouby and Peter Robinson. Real-time inference of complex mental states from facial expressions and head gestures. In *Real-time vision for human-computer interaction*, pages 181–200. Springer, 2005.

[65] Clayton Epp, Michael Lippold, and Regan L Mandryk. Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 715–724. ACM, 2011.

[66] Anita Ferreira and John Atkinson. Designing a feedback component of an intelligent tutoring system for foreign language. *Knowledge-Based Systems*, 22(7):496 – 501, 2009. Artificial Intelligence 2008AI-2008The twenty-eighth {SGAI} International Conference on Artificial Intelligence.

[67] Eugenio Filloy, Luis Puig, and Teresa Rojano. *Educational Algebra: A Theoretical and Empirical Approach*, chapter Algebraic Syntax and Solving Word Problems, pages 141–161. Springer US, Boston, MA, 2008.

[68] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.

[69] Amanda Fleury, Maddy Sugar, and Tom Chau. E-textiles in clinical rehabilitation: a scoping review. *Electronics*, 4(1):173–203, 2015.

[70] Johnny J. R. Fontaine and Klaus R. Scherer. *Components of emotional meaning: A sourcebook*, chapter The global meaning structure of the emotion domain: Investigating the complementarity of multiple perspectives on meaning. Oxford University Press, Oxford, 2013.

[71] Carmen Frank and Elmar Nöth. *Optimizing Eigenfaces by Face Masks for Facial Expression Recognition*, pages 646–654. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[72] W. V. Friesen and P. Ekman. EMFACS-7: Emotional Facial Action Coding System. *Unpublished manuscript, University of California at San Francisco*, 1983.

[73] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, 1901.

[74] Joseph F Grafsgaard, Kristy Elizabeth Boyer, Eric N Wiebe, and James C Lester. Analyzing posture and affect in task-oriented tutoring. In *FLAIRS Conference*, 2012.

[75] Haisong Gu and Qiang Ji. An automated face reader for fatigue detection. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 111–116. IEEE, 2004.

[76] H. Gunes and M. Piccardi. Affect recognition from face and body: early fusion vs. late fusion. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 4, pages 3437–3443 Vol. 4, Oct 2005.

[77] Jaime Gálvez, Eduardo Guzmán, and Ricardo Conejo. A blended e-learning experience in a course of object oriented programming fundamentals. *Knowledge-Based Systems*, 22(4):279 – 286, 2009. Artificial Intelligence (AI) in Blended Learning(AI) in Blended Learning.

[78] Joseph C Hager, Paul Ekman, and Wallace V Friesen. Facial action coding system. *Salt Lake City, UT: A Human Face*, 2002.

[79] Janette R. Hill and Michael J. Hannafin. Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development*, 49(3):37–52, 2001.

[80] Alexander Hong, Yuma Tsuboi, Goldie Nejat, and Beno Benhabib. Affective voice recognition of older adults. *Journal of Medical Devices*, 10(2):020931, 2016.

[81] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[82] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, Jan 1968.

[83] Mary Helen Immordino-Yang and Antonio Damasio. We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, brain, and education*, 1(1):3–10, 2007.

[84] AM Isen. Positive affect and decision making, handbook of emotions, m. *Lewis & J. Haviland-Jones ed*, pages 417–435, 2000.

[85] Gary D James, LS Yee, Gregory A Harshfield, Seymour G Blank, and Thomas G Pickering. The influence of happiness, anger, and anxiety on the blood pressure of borderline hypertensives. *Psychosomatic Medicine*, 48(7):502–508, 1986.

[86] William James. Ii.—what is an emotion? *Mind*, 9(34):188–205, 1884.

[87] Wade Junek. Mind reading: The interactive guide to emotions. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 16(4):182, 2007.

[88] Artūras Kaklauskas, Edmundas Kazimieras Zavadskas, Mark Seniut, Mindaugas Krutinis, Gintautas Dzemyda, Sergėjus Ivanikovas, Voitech Stankevič, Česlovas Šimkevičius, and Aurimas Jaruševičius. Web-based biometric mouse decision support system for user's emotional and labour productivity analysis. *Proceedings of the 25th ISARC*, 2008.

[89] Rana Kaliouby and Peter Robinson. *Affective Computing and Intelligent Interaction: First International Conference, ACII 2005, Beijing, China, October 22-24, 2005. Proceedings*, chapter Generalization of a Vision-Based Computational Model of Mind-Reading, pages 582–589. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[90] Takeo Kanade, Jeffrey F Cohn, and Yingli Tian. Comprehensive database for facial expression analysis. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pages 46–53. IEEE, 2000.

[91] Ashish Kapoor, Winslow Burleson, and Rosalind W Picard. Automatic prediction of frustration. *International journal of human-computer studies*, 65(8):724–736, 2007.

[92] Ashish Kapoor and Rosalind W Picard. Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 677–682. ACM, 2005.

[93] Iftikhar Ahmed Khan, Willem-Paul Brinkman, and Robert Hierons. Towards estimating computer users' mood from interaction behaviour with keyboard and mouse. *Frontiers of Computer Science*, 7(6):943–954, 2013.

[94] Jonghwa Kim and Elisabeth André. Emotion recognition based on physiological changes in music listening. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(12):2067–2083, December 2008.

[95] K. H. Kim, S. W. Bang, and S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42(3):419–427, 2004.

[96] Min-Ki Kim, Miyoung Kim, Eunmi Oh, and Sung-Phil Kim. A review on the computational methods for emotional state estimation from the human EEG. *Comp. Math. Methods in Medicine*, 2013:573734:1–573734:13, 2013.

[97] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, Jan 2012.

[98] Anders Kofod-Petersen. Challenges in case-based reasoning for context awareness in ambient intelligent systems. In *8th European Conference on Case-Based Reasoning, Workshop Proceedings*, pages 287–299, 2006.

[99] A Kolakowska. A review of emotion recognition methods based on keystroke dynamics and mouse movements. In *Human System Interaction (HSI), 2013 The 6th International Conference on*, pages 548–555, June 2013.

[100] Barry Kort, Rob Reilly, and Rosalind W Picard. An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In *icalt*, volume 1, pages 43–47, 2001.

[101] M. B. Kostyunina and M. A. Kulikov. Frequency characteristics of eeg spectra in the emotions. *Neuroscience and Behavioral Physiology*, 26(4):340–343, 1996.

[102] I. Kotsia and I. Pitas. Facial expression recognition in image sequences using geometric deformation features and support vector machines. *Image Processing, IEEE Transactions on*, 16(1):172–187, Jan 2007.

[103] Zirui Lan, Olga Sourina, Lipo Wang, and Yisi Liu. Stability of features in real-time eeg-based emotion recognition algorithm. In *Cyberworlds (CW), 2014 International Conference on*, pages 137–144. IEEE, 2014.

[104] Antonio Lanatà, Antonino Armato, Gaetano Valenza, and Enzo Pasquale Scilingo. Eye tracking and pupil size variation as response to affective stimuli: a preliminary study. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 78–84. IEEE, 2011.

[105] Niels Landwehr, Mark Hall, and Eibe Frank. Logistic model trees. *Mach. Learn.*, 59(1-2):161–205, May 2005.

[106] Peter J Lang, Mark K Greenwald, Margaret M Bradley, and Alfons O Hamm. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3):261–273, 1993.

[107] Mark R Lepper, Maria Woolverton, Donna L Mumme, and J Gurtner. Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. *Computers as cognitive tools*, 1993:75–105, 1993.

[108] Xiaobai Li, Tomas Pfister, Xiaohua Huang, Guoying Zhao, and Matti Pietikäinen. A spontaneous micro-expression database: Inducement, collection and baseline. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–6. IEEE, 2013.

[109] Ying li Tian, Takeo Kanade, and Jeffrey F. Cohn. Recognizing action units for facial expression analysis. *Pattern Analysis and Machine Intelligence*, 2001.

[110] J. J. Lien, T. Kanade, J. F. Cohn, and C. C. Li. Automated facial expression recognition based on facs action units. In *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, FG '98, pages 390–, Washington, DC, USA, 1998. IEEE Computer Society.

[111] James Jenn-Jier Lien, Takeo Kanade, Jeffrey F Cohn, and Ching-Chung Li. Detection, tracking, and classification of action units in facial expression. *Robotics and Autonomous Systems*, 31(3):131–146, 2000.

[112] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen. Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7):1798–1806, July 2010.

[113] Juan-Miguel López-Gil, Jordi Virgili-Gomá, Rosa Gil, and Roberto García. Method for improving eeg based emotion recognition by combining it with synchronized biometric and eye tracking technologies in a non-invasive and low cost way. *Frontiers in Computational Neuroscience*, 10:85, 2016.

[114] F Lotte, M Congedo, A Lécuyer, F Lamarche, and B Arnaldi. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 4(2):R1, 2007.

[115] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE, June 2010.

[116] Simon Lucey, Ahmed Bilal Ashraf, and Jeffrey F Cohn. *Investigating spontaneous facial action recognition through aam representations of the face*. INTECH Open Access Publisher, 2007.

[117] Michael J Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba, and Julien Budynek. The japanese female facial expression (jaffe) database, 1998.

[118] Luis Marco-Giménez, Miguel Arevalillo-Herráez, Francesc J. Ferri, Salvador Moreno-Picot, Jesus Boticario, Olga C. Santos, Sergio Salmeron-Majadas, Mar Saneiro, Raul Uria-Rivas, David Arnau, José Antonio González-Calero, Aladdin Ayesh, Raúl Cabestrero, Pilar Quirós, Pablo Arnau-González, and Naeem Ramzan. Affective and behavioral assessment for adaptive intelligent tutoring systems. In *Late-breaking Results, Posters, Demos, Doctoral Consortium and Workshops Proceedings of the 24th ACM Conference on User Modeling, Adaptation and Personalisation (UMAP 2016), Halifax, Canada, July 13-16, 2016.*, 2016.

[119] Luis Marco-Giménez, Miguel Arevalillo-Herráez, Aladdin Ayesh, and Mariusz Szwoch. An extension to the feedb multimodal database of facial expressions and emotions. *Eurosis ESM 2015*, page 455 – 460, 2015.

[120] Luis Marco-Giménez, Miguel Arevalillo-Herráez, and Cristina Cuhna-Pérez. Eigenexpressions: Emotion recognition using multiple eigenspaces. In *Ib-PRIA*, pages 758–765, 2013.

[121] Sandra P. Marshall. *Schemas in problem solving.* Cambridge University Press, New York, NY, US, 1995.

[122] Derek McColl, Alexander Hong, Naoaki Hatakeyama, Goldie Nejat, and Beno Benhabib. A survey of autonomous human affect detection methods for social robots engaged in natural hri. *Journal of Intelligent & Robotic Systems*, 82(1):101–133, 2016.

[123] Rollin McCraty, Mike Atkinson, William A. Tiller, Glen Rein, and Alan D. Watkins. The effects of emotions on short-term power spectrum analysis of heart rate variability. *The American Journal of Cardiology*, 76(14):1089 – 1093, 1995.

[124] Bethany T McDaniel, Sidney D'Mello, Brandon G King, Patrick Chipman, Kristy Tapp, and AC Graesser. Facial features for affective state detection in learning environments. In *Proceedings of the 29th Annual Cognitive Science Society*, pages 467–472. Citeseer, 2007.

[125] Richard A. McFarland. Relationship of skin temperature changes to the emotions accompanying music. *Biofeedback and Self-regulation*, 10(3):255–267, 1985.

[126] G. McKeown, M. F. Valstar, R. Cowie, and M. Pantic. The semaine corpus of emotionally coloured character interactions. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pages 1079–1084, July 2010.

[127] Albert Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4):261–292, 1996.

[128] Albert Mehrabian and James A Russell. *An approach to environmental psychology.* the MIT Press, 1974.

[129] Bo Melin and Ulf Lundberg. A biopsychosocial approach to work-stress and musculosketal disorders. *Journal of Psychophysiology*, 1997.

[130] Marvin Minsky and Seymour Papert. Perceptrons cambridge, 1969.

[131] S. Mota and R.W. Picard. Automated posture analysis for detecting learner's interest level. In *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03. Conference on*, volume 5, pages 49–49, June 2003.

[132] Subhas Chandra Mukhopadhyay. Wearable sensors for human activity monitoring: A review. *IEEE Sensors Journal*, 15(3):1321–1330, 2015.

[133] Arturo Nakasone, Helmut Prendinger, and Mitsuru Ishizuka. Emotion recognition from electromyography and skin conductance. In *Proc. of the 5th International Workshop on Biosignal Interpretation*, pages 219–222. Citeseer, 2005.

[134] Costanza Navarretta. Predicting emotions in facial expressions from the annotations in naturally occurring first encounters. *Knowledge-Based Systems*, 71:34–40, 2014.

[135] M. Pantic and I. Patras. Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 36(2):433–449, April 2006.

[136] M. Pantic and L.J.M. Rothkrantz. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE*, 91(9):1370–1390, Sept 2003.

[137] Maja Pantic and Marian Stewart Bartlett. *Machine analysis of facial expressions.* I-Tech Education and Publishing, 2007.

[138] Maja Pantic and Leon J. M. Rothkrantz. Facial action recognition for facial expression analysis from static face images, 2004.

[139] Maja Pantic, Michel Valstar, Ron Rademaker, and Ludo Maat. Web-based database for facial expression analysis. In *2005 IEEE international conference on multimedia and Expo*, pages 5–pp. IEEE, 2005.

[140] Luc Paquette, Jonathan Rowe, Ryan Baker, Bradford Mott, James Lester, Jeanine DeFalco, Keith Brawner, Robert Sottilare, and Vasiliki Georgoulas. Sensor-free or sensor-full: A comparison of data modalities in multi-channel affect detection. *International Educational Data Mining Society*, 2016.

[141] Zachary A. Pardos, Ryan S. J. D. Baker, Maria O. C. Z. San Pedro, Sujith M. Gowda, and Supreeth M. Gowda. Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, LAK '13, pages 117–124, New York, NY, USA, 2013. ACM.

[142] Reinhard Pekrun, Thomas Goetz, Lia M Daniels, Robert H Stupnisky, and Raymond P Perry. Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, 102(3):531, 2010.

[143] R. W. Picard. Affective computing, 1995.

[144] Senya Polikovsky, Yoshinari Kameda, and Yuichi Ohta. Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. In *Crime Detection and Prevention (ICDP 2009), 3rd International Conference on*, pages 1–6. IET, 2009.

[145] Luis Puig. Researching (algebraic) problem solving from the perspective of local theoretical models. *Procedia - Social and Behavioral Sciences*, 8:3 – 16, 2010. International Conference on Mathematics Education Research 2010 (ICMER 2010).

[146] J. R. Quinlan. Induction of decision trees. *Mach. Learn.*, 1(1):81–106, March 1986.

[147] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[148] Pierre Rainville, Antoine Bechara, Nasir Naqvi, and Antonio R. Damasio. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology*, 61(1):5 – 18, 2006. Psychophiosology and Cognitive Neuroscience.

[149] Matthias Reif, Faisal Shafait, Markus Goldstein, Thomas Breuel, and Andreas Dengel. Automatic classifier selection for non-experts. *Pattern Analysis and Applications*, 17(1):83–96, 2014.

[150] Mary S. Riley, James G. Greeno, Joan I. Heller, and National Institute of Education (U.S.). *Development of children's problem-solving ability in arithmetic [microform] / Mary S. Riley, James G. Greeno, Joan I. Heller.* Learning Research and Development Center, University of Pittsburgh [Pittsburgh, Pa.], 1984.

[151] Sara E Rimm-Kaufman and Jerome Kagan. The psychological significance of changes in skin temperature. *Motivation and Emotion*, 20(1):63–78, 1996.

[152] Frank Rosenblatt. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, DTIC Document, 1961.

[153] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980.

[154] James A Russell, Jo-Anne Bachorowski, and Jose-Miguel Fernandez-Dols. Facial and vocal expressions of emotion. *Annual review of psychology*, 54(1):329–349, 2003.

[155] Jennifer Sabourin, Bradford Mott, and James C Lester. Modeling learner affect with theoretically grounded dynamic bayesian networks. In *International Conference on Affective Computing and Intelligent Interaction*, pages 286–295. Springer, 2011.

[156] Hiroaki Sakoe. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26:43–49, 1978.

[157] Sergio Salmeron-Majadas, Olga C Santos, and Jesus G Boticario. An evaluation of mouse and keyboard interaction indicators towards non-intrusive and low cost affective modeling in an educational context. *Procedia Computer Science*, 35:691–700, 2014.

[158] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.*, 11(5):561–580, October 2007.

[159] David Sander, Didier Grandjean, and Klaus R. Scherer. 2005 special issue: A systems approach to appraisal mechanisms in emotion. *Neural Netw.*, 18(4):317–352, May 2005.

[160] Olga C Santos. Emotions and personality in adaptive e-learning systems: an affective computing perspective. In *Emotions and Personality in Personalized Services*, pages 263–285. Springer, 2016.

[161] Olga C Santos, Sergio Salmeron-Majadas, and Jesus G Boticario. Emotions detection from math exercises by combining several data sources. In *International Conference on Artificial Intelligence in Education*, pages 742–745. Springer, 2013.

[162] Olga C Santos, Mar Saneiro, Sergio Salmeron-Majadas, and Jesus G Boticario. A methodological approach to eliciting affective educational recommendations. In *2014 IEEE 14th International Conference on Advanced Learning Technologies*, pages 529–533. IEEE, 2014.

[163] Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(6):1113–1133, 2015.

[164] Arman Savran, Koray Ciftci, Guillaume Chanel, Javier Mota, Luong Hong Viet, Blent Sankur, Lale Akarun, Alice Caplier, and Michele Rombaut. Emotion detection in the loop from brain signals and facial images. In *Proceedings of the eNTERFACE 2006 Workshop*. Dubrovnik (Croatia), July 2006.

[165] Nicu Sebe, Ira Cohen, Theo Gevers, and Thomas S. Huang. Multimodal approaches for emotion recognition: a survey. In Simone Santini, Raimondo Schettini, and Theo Gevers, editors, *Internet Imaging VI: Proceedings of*

the Society of Photo-Optical Instrumentation Engineers, volume 5670, pages 56–67. SPIE, 2005.

[166] Nicu Sebe, Michael S Lew, Yafei Sun, Ira Cohen, Theo Gevers, and Thomas S Huang. Authentic facial expression analysis. *Image and Vision Computing*, 25(12):1856–1863, 2007.

[167] Matthew Shreve, Sridhar Godavarthy, Dmitry Goldgof, and Sudeep Sarkar. Macro-and micro-expression spotting in long videos using spatio-temporal strain. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 51–56. IEEE, 2011.

[168] Matthew Shreve, Sridhar Godavarthy, Vasant Manohar, Dmitry Goldgof, and Sudeep Sarkar. Towards macro-and micro-expression spotting in video using strain patterns. In *Applications of Computer Vision (WACV), 2009 Workshop on*, pages 1–6. IEEE, 2009.

[169] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic. A multimodal database for affect recognition and implicit tagging. *Affective Computing, IEEE Transactions on*, 3(1):42–55, Jan 2012.

[170] Mohammad Soleymani, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, 7(1):17–28, 2016.

[171] Robert A Sottilare, Arthur Graesser, Xiangen Hu, and Heather Holden. *Design Recommendations for Intelligent Tutoring Systems: Volume 1-Learner Modeling*, volume 1, p. ii. US Army Research Laboratory, 2013.

[172] Robert A Sottilare and Michael D Proctor. Passively classifying student mood and performance within intelligent tutors. *Educational Technology & Society*, 15(2):101–114, 2012.

[173] Kent A. Spackman. Signal detection theory: Valuable tools for evaluating inductive learning. In *Proceedings of the Sixth International Workshop on Machine Learning*, pages 160–163, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc.

[174] Steven K. Sutton and Richard J. Davidson. Prefrontal brain asymmetry: A biological substrate of the behavioral approach and inhibition systems. *Psychological Science*, 8(3):204–210, 1997.

[175] M. Szwoch. Feedb: A multimodal database of facial expressions and emotions. In *Human System Interaction (HSI), 2013 The 6th International Conference on*, pages 524–531, June 2013.

[176] Mariusz Szwoch. On facial expressions and emotions RGB-D database. In Stanislaw Kozielski, Dariusz Mrozek, Pawel Kasprowski, Bozena Malysiak-Mrozek, and Daniel Kostrzewa, editors, *Beyond Databases, Architectures, and Structures*, volume 424 of *Communications in Computer and Information Science*, pages 384–394. Springer International Publishing, 2014.

[177] Wioleta Szwoch. Emotion recognition using physiological signals. In *Proceedings of the Mulitimedia, Interaction, Design and Innnovation, MIDI 2015, Warsaw, Poland, June 29-30, 2015*, pages 15:1–15:8, 2015.

[178] Hao Tang and Thomas S Huang. 3d facial expression recognition based on properties of line segments connecting facial feature points. In *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008.

[179] Ying-li Tian, Takeo Kanade, and Jeffrey F Cohn. Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 229–234. IEEE, 2002.

[180] Yingli Tian, Takeo Kanade, and Jeffrey F Cohn. Facial expression recognition. In *Handbook of face recognition*, pages 487–519. Springer, 2011.

[181] Yan Tong, Jixu Chen, and Qiang Ji. A unified probabilistic framework for spontaneous facial action modeling and understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(2):258–273, 2010.

[182] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, January 1991.

[183] Michel F Valstar, Maja Pantic, Zara Ambadar, and Jeffrey F Cohn. Spontaneous vs. posed facial behavior: automatic analysis of brow actions. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 162–170. ACM, 2006.

[184] Paul Viola and Michael J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, May 2004.

[185] Lisa M Vizer, Lina Zhou, and Andrew Sears. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies*, 67(10):870–886, 2009.

[186] J. Wagner, Jonghwa Kim, and E. Andre. From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 940–943, July 2005.

[187] Zhen Wen and Thomas S Huang. Capturing subtle facial motions in 3d face tracking. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1343–1350. IEEE, 2003.

[188] Jacob Whitehill, Zewelanji Serpell, Yi-Ching Lin, Aysha Foster, and Javier R Movellan. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. *IEEE Transactions on Affective Computing*, 5(1):86–98, 2014.

[189] Qi Wu, Xunbing Shen, and Xiaolan Fu. The machine knows what you are hiding: an automatic micro-expression recognition system. In *International Conference on Affective Computing and Intelligent Interaction*, pages 152–162. Springer, 2011.

[190] Zhongzhe Xiao, Emmanuel Dellandréa, Weibei Dou, and Liming Chen. Classification of Emotional Speech Based on an Automatically Elaborated Hierarchical Classifier. *ISRN Signal Processing*, 2011.

[191] Mohammed Yeasin, Baptiste Bullot, and Rajeev Sharma. Recognition of facial expressions and measurement of levels of interest from video. *IEEE Transactions on Multimedia*, 8(3):500–508, 2006.

[192] Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, Jan 2009.

[193] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015.

[194] Wei-Long Zheng, Jia-Yi Zhu, and Bao-Liang Lu. Identifying stable patterns over time for emotion recognition from EEG. *CoRR*, abs/1601.02197, 2016.

[195] Philippe Zimmermann, Sissel Guttormsen, Brigitta Danuser, and Patrick Gomez. Affective computing—a rationale for measuring mood with mouse and keyboard. *International journal of occupational safety and ergonomics*, 9(4):539–551, 2003.