



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Departamento de Sistemas Informáticos y Computación

Universitat Politècnica de València

Una primera Aproximación hacia la Computación Afectiva en Entornos de Realidad Virtual Multi-Modales e Interactivos

MÁSTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL,
RECONOCIMIENTO DE FORMAS E IMAGEN DIGITAL

TRABAJO FIN DE MÁSTER

Curso: 2021-2022

Autor:

Juan Antonio De Rus Arance

Tutores:

Carlos Carrascosa Casamayor

Mario Montagud Climent

Reconocimientos

Este trabajo ha sido posible gracias a contar con material cedido por la Universidad de Valencia como sensores, placas electrónicas y otros dispositivos como el Head Mounted Display (HMD) Oculus Quest.

Quiero agradecer tanto a la Universidad de Valencia como a la Fundación I2CAT por concedernos permiso para utilizar escenarios virtuales desarrollados en distintos proyectos en los que ha colaborado el autor del presente documento.

Agradezco también a la Universidad Politécnica de Valencia y al tribunal que juzgará este trabajo por acceder a adelantar la convocatoria de defensa.

Quiero dar las gracias también a mi tutor en este trabajo, Carlos Carrascosa Casamayor, por su guía y apoyo.

Finalmente, agradezco a quien en esta ocasión es mi tutor externo, Mario Montagud Climent, por estar siempre pendiente y por ayudarme siempre que lo necesito desde hace bastantes años y espero que durante bastantes más.

Resumen

La computación afectiva es un campo de la informática con muchas aplicaciones por desarrollar y explotar. En este trabajo aplicaremos la computación afectiva a entornos de Realidad Virtual (RV) interactivos para estudiar la respuesta emotiva de los usuarios a distintos estímulos.

En primer lugar, se ha creado un dataset propio, recopilando los datos fisiológicos de distintos usuarios tras exponerlos a distintos estímulos con tal de provocarles emociones, junto con sus respuestas a cuestionarios del tipo “Self-Assessment Manikin” para inferir las emociones sentidas. Estos datos fueron recopilados finalmente con un prototipo creado con una placa Arduino y varios sensores conectados y programados.

Dicho dataset se ha utilizado posteriormente para para crear un modelo de regresión de las emociones sentidas por cada usuario usando una estructura de redes neuronales LSTM (Long Short-Term Memory), y se ha aplicado para la observación de la respuesta emotiva a distintos estímulos en escenarios de RV interactivos que se han preparado. Entre los estímulos comparados en los escenarios de RV están el uso de audio máquina o el uso de audio humano, el uso de distintos tipos de subtítulos, y el uso de texto o de audio para describir puntos de interés.

En cuanto a los resultados, los estímulos seleccionados no han provocado una gran respuesta emotiva por parte del usuario. Por otra parte, el modelo de regresión ha tenido resultados aceptables a la hora de estimar la respuesta emotiva de los usuarios en base a sus métricas fisiológicas.

Se espera que este estudio preliminar abra la puerta a una nueva línea de investigación en esta área, materializándose en una Tesis Doctoral. Los resultados obtenidos no han sido conclusivos por falta de medios, como el: reducido número de voluntarios para el estudio, y baja calidad de los sensores utilizados para recopilación de métricas (a falta de acceso a otros mejores), así como por las limitaciones de tiempo.

Palabras Clave: Computación Afectiva, Realidad Virtual, Sensores, LSTM, Inteligencia Artificial.

Resum

La computació afectiva és un camp de la informàtica amb moltes aplicacions per desenvolupar i explotar. En aquest treball aplicarem la computació afectiva a entorns de Realitat Virtual (RV) interactius per estudiar la resposta emotiva dels usuaris a diferents estímuls.

En primer lloc, s'ha creat un dataset propi, recopilant les dades fisiològiques de diferents usuaris, després d'exposar-los a diferents estímuls per provocar-los emocions, junt amb les seues respostes a qüestionaris del tipus "Self-Assessment Manikin" per inferir les emocions sentides. Aquestes dades van ser recopilades finalment amb un prototip creat amb una placa Arduino i diversos sensors connectats i programats

Aquest dataset s'ha utilitzat posteriorment per a crear un model de regressió de les emocions sentides usant una estructura de xarxes neuronals LSTM (Long Short-Term Memory), i s'ha aplicat per a l'observació de la resposta emotiva a cada estímulo en els escenaris de RV que s'han preparat. Entre els estímuls comparats en els escenaris de RV estan l'ús d'àudio màquina o l'ús d'àudio humà, l'ús de diferents tipus de subtítols, i l'ús de text o d'àudio per a descriure punts d'interès.

Quant als resultats, els estímuls seleccionats no han provocat una gran resposta emotiva per part de l'usuari. D'altra banda el model de regressió ha tingut resultats acceptables a l'hora d'estimar la resposta emotiva dels usuaris sobre la base de les seues mètriques fisiològiques.

S'espera que aquest estudi preliminar siga el punt de partida a una nova línia d'investigació en aquesta àrea, materialitzant-se en una Tesi Doctoral. Els resultats obtinguts no han sigut conclusius per falta de mitjans, com un reduït nombre de voluntaris per a l'estudi, i baixa qualitat dels sensors utilitzats per a recopilació de mètrica (a falta d'accés a altres millors), així com per les limitacions temporals.

Paraules Clau: Computació Afectiva, Realitat Virtual, Sensors, LSTM, Intel·ligència Artificial.

Abstract

Affective computing is a field of computing with many applications to develop and exploit. In this work we will apply affective computing to interactive Virtual Reality (VR) environments to study the emotional response of users to different stimuli.

First, a dataset has been created, by collecting the physiological data from different users after exposing them to different stimuli to provoke emotions, together with their responses to “Self-Assessment Manikin” questionnaires to infer the emotions felt. This data was finally collected with a prototype created with an Arduino board and several sensors connected and programmed.

This dataset has subsequently been used to create a regression model of the emotions felt by each user using an LSTM (Long Short-Term Memory) neural network structure, and it has been applied to observe the emotional response to each stimulus in the VR scenarios that have been prepared and presented to the users. Among the stimuli compared in VR scenarios are the use of computer-generated audio or the use of human audio, the use of different types of subtitles, and the use of text or audio to describe points of interest.

Regarding the results, the selected stimuli have not elicited a great emotional response from the user. On the other hand, the regression model has had acceptable results when estimating the emotional response of users based on their physiological metrics.

This preliminary study is expected to open the door to a new research line in this field, being further developed in a PhD Thesis. The obtained results have not been conclusive due to lack of means, like reduced number of volunteers for the study and low quality of the sensors used to collect the metrics (in the absence of access to better ones), as well as time constraints.

Keywords: Affective Computing, Virtual Reality, Sensors, LSTM, Artificial Intelligence.

Tabla de contenido

Reconocimientos.....	I
Resumen	II
Abstract.....	IV
Tabla de contenido.....	V
Lista de Ilustraciones.....	VIII
Lista de tablas.....	IX
Lista de ecuaciones	IX
Lista Acrónimos y abreviaturas	X
1. Introducción	1
1.1. Descripción del problema.....	1
1.2. Objetivos	1
1.3. Estructura de la memoria	2
2. Estado del arte	3
2.1. Entornos de realidad virtual multimodales e interactivos.....	3
2.2. Computación afectiva.....	4
2.2.1. Aplicaciones actuales.....	4
2.2.2. Aplicaciones Futuras Posibles	5
2.3. Representaciones de la respuesta emotiva	6
2.3.1. Circumplex model.....	6
2.3.2. Vector Model.....	7
2.3.3. Pana model.....	7
2.3.4. Plutchik's model	8
2.3.5. PAD Model.....	9
2.4. Estudios en el campo de la computación afectiva relacionados.....	9
2.4.1. AffectiveROAD.....	9
2.4.2. Amigos.....	9
2.4.3. CASE.....	10
2.4.4. Cognitive Load Inference	10
2.4.5. DEAP.....	10
2.4.6. Dreamer.....	10
2.4.7. MAHNOB	11
2.4.8. Resumen de datasets de estudios similares	11

2.5.	Estudios relacionados que integren computación afectiva con entornos de RV multimodales.	11
2.5.1.	Estudio de la respuesta emotiva en escenarios de RV	11
2.5.2.	Uso de computación afectiva combinada con escenarios de RV como herramienta de apoyo al aprendizaje	12
2.5.3.	Uso de computación afectiva combinada con escenarios de RV como herramienta de rehabilitación	12
2.6.	Modelo LSTM de red neuronal	13
2.6.1.	Redes neuronales	13
2.6.2.	Redes neuronales recurrentes	13
2.6.3.	Modelo LSTM.....	14
3.	Experimentación	18
3.1.	Librerías.....	18
3.1.1.	TensorFlow y Keras.....	18
3.1.2.	Pandas y Numpy.....	18
3.1.3.	Scipy y Sklearn	18
3.1.4.	Librerías Adicionales.....	18
3.2.	Dispositivos de captura de datos fisiológicos consultados	18
3.2.1.	Pulseras inteligentes.....	19
3.2.2.	MySignals.....	20
3.2.3.	Sensores seleccionados y utilizados	20
3.3.	Entornos de Realidad Virtual Creados y Utilizados	21
3.4.	Creación de datasets	24
3.4.1.	Selección de videos para evocar emociones.....	24
3.4.2.	Cuestionarios.....	29
3.4.3.	Metodología de las pruebas	31
3.4.4.	Pruebas en entornos de RV	31
3.5.	Exploración y Preprocesado de los datos	32
3.5.1.	Exploración	32
3.5.2.	Preprocesado.....	33
3.6.	Creación y entrenamiento del modelo LSTM	34
3.6.1.	Métricas utilizadas en experimentos	35
3.6.2.	División del dataset para realizar experimentos.....	36
3.6.3.	Experimentos.....	37
3.6.4.	Modelo Final.....	41
4.	Resultados.....	43
4.1.	Creación del dataset.....	43

4.2.	Resultados aplicados a Escenarios de RV	44
4.3.	Resultados del modelo de predicción de emociones	47
5.	Conclusiones y Trabajo Futuro	49
5.1.	Análisis de los resultados.....	49
5.1.1.	Calidad de los datos.....	49
5.1.2.	Tests en escenarios de RV	50
5.2.	Conclusiones.....	51
5.3.	Problemas encontrados durante la realización del trabajo	52
5.4.	Trabajo futuro	52
	Bibliografía.....	54

Lista de Ilustraciones

Ilustración 2.1 - Ejemplo de HMD y Controladores en entornos de RV	4
Ilustración 2.2 - Guantes Hápticos.....	4
Ilustración 2.3 - Cintas Omnidireccionales para caminar en entornos de RV	4
Ilustración 2.4 - Modelo Circunflejo de Russell con emociones principales	7
Ilustración 2.5 - Modelo Circunflejo de Russell por regiones	7
Ilustración 2.6 - Modelo Bidireccional Vector Model	7
Ilustración 2.7 - Modelo PANA	8
Ilustración 2.8 - Modelo PANA representando como modelo circunflejo girado 45º	8
Ilustración 2.9 - Diadas del modelo de Plutchik	8
Ilustración 2.10 - Modelo de Plutchik.....	8
Ilustración 2.11 - Áreas que ocupan distintas emociones en modelo PAD.....	9
Ilustración 2.12 - Puntos principales de distintas emociones en modelo PAD	9
Ilustración 2.13 - Estructura de neurona artificial simple.....	13
Ilustración 2.14 - Estructura típica de red neuronal	13
Ilustración 2.15 - Red Neuronal Recurrente	14
Ilustración 2.16 - Comparación simplificada de distintas arquitecturas de redes neuronales ...	14
Ilustración 2.17 - Diferencias RNN y LSTM	15
Ilustración 2.18 - Esquema LSTM	15
Ilustración 3.1 - Kit de sensores "MySignals SW" desarrollado por la empresa Liberia.....	20
Ilustración 3.2 - Placa electrónica Arduino Nano	20
Ilustración 3.3 - Sensor "Max30102" para la medición de la frecuencia cardiaca de la empresa Maxim Integrated.....	21
Ilustración 3.4 - Sensor GSR de la empresa Sichiray.....	21
Ilustración 3.5 - Capturas de los teatros reconstruidos: (izquierda) Olivera; (derecha) Globe ...	22
Ilustración 3.6 - Ejemplo de Punto de Interés con Texto	22
Ilustración 3.7 - Capturas de diferentes pantallas y opciones de menú	22
Ilustración 3.8 - Captura de obras en sala de croma, con vídeo inmersivo 180º estereoscópico y audio espacial.....	23
Ilustración 3.9 - Vista desde balcón del escenario de RV del Corral de Comedias de la Olivera .	23
Ilustración 3.10 - Obra con subtítulos con estilo	24
Ilustración 3.11 - Video de agresión gratuita a pareja.....	25
Ilustración 3.12 - Sacrificio de Cerdos	25
Ilustración 3.13 - Mongo Cat Espacial	26
Ilustración 3.14 - Corto animado "The Crow and the Raven"	26
Ilustración 3.15 - Imágenes utilizadas en el Self Assessment Manikin questionnaire para medir valores de Arousal, Valence y Dominance.....	30
Ilustración 3.16 - Uso de EmojiGrid para evaluar emociones en escenarios de RV	30
Ilustración 3.17 - Distribución de los videos por cuadrante emocional	33
Ilustración 3.18 - Valores de Valence originales e interpolados	34
Ilustración 3.19 - Matriz de confusión	35
Ilustración 3.20 - Distancia entre predicciones y emociones reales. Tests del usuario 3.	36
Ilustración 3.21 - Partición de dataset mezclando videos y usuarios	37
Ilustración 3.22 - Resultados clasificación último experimento	40
Ilustración 3.23 - Resultados en distancia euclídea último experimento	41
Ilustración 3.24 - Estructura modelo final entrenado.....	42
Ilustración 4.1 - Emociones provocadas por los videos en los usuarios	43

Ilustración 4.2 - Emociones provocadas por los tests en los usuarios	44
Ilustración 4.3 Resultados Reales - Test 1: Obra con audio máquina y texto plano	45
Ilustración 4.4 Resultados Reales - Test 1: Obra con audio humano y texto plano	45
Ilustración 4.5 Resultados Reales - Test 1: Obra con audio humano y texto con estilo.....	46
Ilustración 4.7 Resultados Reales - Test 2: Punto de Interés con Texto	46
Ilustración 4.7 Resultados Reales - Test 2: Punto de Interés con Audio	46
Ilustración 4.8 - Resultados de los tests en escenarios de RV por usuario y en conjunto	47
Ilustración 5.1 - Valores de GSR	49

Lista de tablas

Tabla 2.1 - Datasets de Estudios Similares	11
Tabla 2.2 - Métricas Clave de datasets Existentes	11
Tabla 3.1 - Pulseras Smartwatch Consultadas	19
Tabla 3.2 - Listado de videos utilizados en pruebas	29
Tabla 3.3 - Resultados sobre el subconjunto de validación al realizar experimentos.....	39
Tabla 4.1 - Resumen resultados de los tests en escenarios de RV	48

Lista de ecuaciones

Ecuación 2.1 - Ecuación Funcionamiento LSTM 1.....	16
Ecuación 2.2 - Ecuación Funcionamiento LSTM 2.....	16
Ecuación 2.3 - Ecuación Funcionamiento LSTM 3.....	16
Ecuación 2.4 - Ecuación Funcionamiento LSTM 4.....	16
Ecuación 2.5 - Ecuación Funcionamiento LSTM 5.....	17
Ecuación 2.6 - Ecuación Funcionamiento LSTM 6.....	17

Lista Acrónimos y abreviaturas

A continuación, listamos los acrónimos que utilizaremos en este documento.

ACC	Acceleration: Aceleración en distintos ejes
BR	Breath Rate: Ritmo de respiración
BVP	Blood Volume Pulse: Volumen del Pulso Sanguíneo extraído con un fotopleletismógrafo
CASE	Continuously Annotated Signals of Emotion. Dataset de estudio relacionado.
DEAP	Database for Emotion Analysis using Physiological Signals. Dataset de estudio relacionado.
ECG	Electrocardiography: Electrocardiograma. A partir del ECG se puede sacar la frecuencia cardíaca.
EDA	Electrodermic activity: Actividad electrodérmica en microsiemens (μS)
EEG	Electroencephalogram: Electroencefalograma. Es una prueba para detectar la actividad eléctrica del cerebro.
EMG	Electromyography: Electromiografía. Mide la actividad eléctrica producida por los músculos esqueléticos
EULA	End-User License Agreement: Acuerdo de licencia de usuario final
GSR	Galvanic Skin Response: Respuesta galvánica de la piel. Medida en Ohms o en microsiemens (μS). Distintos datasets pueden medirlo de forma distinta con sensores entre dedos de la misma mano, o entre el pie y la mano.
HMD	Head Mounted Display (la traducción literal al castellano es “Pantalla montada en la cabeza”), consiste en un dispositivo de visualización parecido a un casco. Permite reproducir imágenes creadas por computador en una pantalla próxima a los ojos.
HR	Heart Rate: Frecuencia Cardíaca
HVHA	High Valence High Arousal. En el modelo circunflejo de representación de las emociones de Russell es el cuadrante donde la Valencia y el Arousal son positivos. Se corresponde con emociones como el regocijo, la excitación o la felicidad.
HVLA	High Valence Low Arousal. En el modelo circunflejo de representación de las emociones de Russell es el cuadrante donde la Valencia es positiva y el Arousal negativo. Se corresponde con emociones como la calma.
LSTM	Long Short-Term Memory. Es un tipo de red neuronal recurrente.
LVHA	Low Valence High Arousal. En el modelo circunflejo de representación de las emociones de Russell es el cuadrante donde la Valencia es negativa y el Arousal positivo. Se corresponde con emociones como la ansiedad, el enojo o el miedo.

LVLA	Low Valence Low Arousal. En el modelo circunflejo de representación de las emociones de Russell es el cuadrante donde la Valencia y el Arousal son negativas. Se corresponde con emociones como el aburrimiento o la tristeza.
MAE	Mean Absolute Error
MSE	Mean Squared Error
Resp	Respiración calculada a partir de la expansión de la cavidad torácica.
RNN	Recurrent Neural Network : Red Neuronal Recurrente
RV	Realidad Virtual
SAM	Self-Assessment manikin. Es un tipo de cuestionario utilizado para calificar emociones utilizando dibujos.
SVM	Support Vector Machines: máquinas de vectores de soporte. Es un tipo de herramienta utilizada para tareas de clasificación y regresión.
Temp	Temperatura en escala Celsius (°C)

1. Introducción

En este capítulo describiremos el problema que nos planteamos, definimos los objetivos que nos proponemos y describimos la estructura que seguiremos en esta memoria.

1.1. Descripción del problema

La computación afectiva es un campo con mucho potencial de mejora que se centra en el uso de inteligencia artificial para reconocer emociones e interpretarlas con distintos objetivos. Ésta combinada con el uso la realidad virtual se podría aplicar a muchas áreas distintas: educación remota adaptada a las necesidades de cada estudiante, visitas virtuales personalizadas a lugares lejanos, apoyo a personas con autismo, experiencias virtuales que evolucionan y se adaptan a lo que siente el usuario en cada momento, etc.

En este trabajo realizaremos una primera aproximación a la aplicación de la Computación Afectiva en entornos virtuales multimodales e interactivos, esto es, usando dichos entornos estudiaremos la respuesta afectiva o emocional que generan en los usuarios utilizando distintas técnicas para evaluar su respuesta. Intentamos con esto abrir una nueva línea de investigación que se materialice en un doctorado.

Evaluaremos distintas métricas, parámetros y sensores que podemos utilizar para la medición de la computación afectiva acorde a distintos criterios como la extensión de uso de determinados dispositivos, la escalabilidad, la factibilidad o los resultados obtenidos en estudios anteriores. Asimismo, analizaremos la selección y adopción de modelos de clasificación de los datos obtenidos con las métricas anteriores para computar dicha respuesta afectiva de manera efectiva.

Pretendemos hacer un modelo escalable que funcione con métricas fisiológicas medibles con sensores comunes y asequibles, como por ejemplo aquellos presentes en relojes y pulseras inteligentes, por lo que intentaremos no utilizar un elevado número de sensores, ni sensores que sean intrusivos o muy especializados o caros.

El trabajo realizado está relacionado con el máster cursado en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital, pues trabajamos con escenarios de realidad virtual creados por el autor y usamos herramientas de inteligencia artificial y machine learning. Además, el autor de este trabajo ha sido co-autor de 1 artículo sobre la medición de métricas en aplicaciones multimedia [46] y 2 artículos sobre el diseño de escenarios de RV multi-modales con estrategias de presentación de estímulos [47] [45].

1.2. Objetivos

Nuestro objetivo principal en este trabajo es evaluar qué eventos y estímulos producen una mayor respuesta emotiva en los usuarios en escenarios de Realidad Virtual (RV), así como determinar en cierto modo cómo se sienten los usuarios en estos entornos. Además, entrenar modelos capaces de predecir dicha respuesta emotiva a partir de datos fisiológicos.

A parte del objetivo principal, se han definido una serie de objetivos específicos en este trabajo, que se listan a continuación:

- Creación de escenarios RV incluyendo distintos eventos y estímulos multimodales para poder estudiar posteriormente la respuesta emotiva de los usuarios cuando son sometidos a distintos eventos/estímulos.
- Análisis y selección de métricas y sensores para la recolección de señales fisiológicas.

- Creación de datasets para la creación de los modelos de inferencia. Para ello recopilaremos datos fisiológicos de los usuarios a la vez que les hacemos visionar un conjunto de materiales audiovisuales (videos) de distinta índole para provocar distintas emociones en los mismos. Posteriormente, mediante el uso de cuestionarios específicos para dicha tarea, extraeremos la emoción principal sentida por cada usuario durante la visualización de cada video y crearemos un pequeño dataset que relacionará datos fisiológicos con emociones experimentadas por los usuarios.
- Investigar sobre datasets semejantes y si es posible usarlos para expandir nuestro propio dataset.
- Creación de un modelo de predicción a partir del dataset creado, intentando maximizar la precisión en la medida de lo posible.
- Realización de pruebas en los escenarios RV creados, para los que, al igual que durante la creación de datasets, someteremos al usuario a distintos estímulos y eventos a la vez que recopilamos datos fisiológicos. Posteriormente, realizaremos otro cuestionario al usuario para inferir sus emociones. En este caso, estos datos servirán para calcular/estimar la respuesta emotiva.
- Aplicar modelo a los datos recopilados durante las pruebas en escenarios RV para inferir qué eventos provocan una respuesta emotiva mayor y más satisfactoria en usuarios.

1.3. Estructura de la memoria

Este documento está estructurado en 5 capítulos.

En el primer capítulo describimos la tarea que nos hemos propuesto realizar, en qué se relaciona con nuestro perfil, y enumeramos los distintos objetivos propuestos por completar.

A continuación, en el segundo capítulo hacemos una revisión del estado del arte en el ámbito de la RV y la computación afectiva. Indagamos en los distintos modelos de representación y catalogación de las emociones que existen, y consultamos estudios relacionados con nuestro trabajo. Además, hacemos una breve explicación de las herramientas que utilizaremos en el trabajo, las redes neuronales LSTM.

El tercer capítulo es el más extenso: detallamos los sensores de captura de datos seleccionados, presentamos los entornos de RV creados para nuestro trabajo, y describimos el proceso de creación de nuestro dataset que hemos seguido. A partir del dataset creado, vemos como hemos tratado los datos y a continuación vemos el proceso de creación de nuestro modelo LSTM final con todos los experimentos realizados hasta completarlo.

En el cuarto capítulo estudiamos los resultados obtenidos.

Finalmente, en el capítulo quinto presentamos nuestros análisis y conclusiones, y detallamos líneas de posible trabajo futuro.

2. Estado del arte

En este capítulo realizaremos una pequeña introducción donde explicamos en qué consisten los entornos de RV, una introducción a la computación afectiva, explicamos distintos modelos utilizados para la representación de emociones, consultamos distintos trabajos en el campo de la computación afectiva para el estudio de las emociones utilizando herramientas de inteligencia artificial y machine learning, y finalmente consultamos trabajos que integren los campos de la computación afectiva y los entornos de realidad virtual.

2.1. Entornos de realidad virtual multimodales e interactivos

Los entornos de RV son entornos/escenarios virtuales generados por computador pensados para ser experimentados por un usuario haciendo uso de distintos dispositivos de visualización, preferentemente HMD (Head Mounted Display), que consisten en visores colocados en la cabeza por los que mira el usuario y que aportan una mayor sensación de “inmersión” en el entorno virtual. En la Ilustración 2.1 podemos ver un ejemplo de HMD y controladores. Esta inmersión se consigue gracias a distintas características de los HMD como:

- Ofrecen una imagen distinta a cada ojo, dando la sensación de una visión estereoscópica semejante a la de la visión natural al mirar a objetos en tres dimensiones. Evita la sensación de estar mirando a una pantalla plana.
- Registran movimientos del usuario y pueden usar dichos datos para ejecutar un movimiento semejante en el escenario virtual. Por ejemplo, si un usuario con un HMD colocado gira el cuello hacia la derecha, el movimiento es recogido y la cámara en el escenario virtual gira a la vez. Esto hace innecesario el uso de otros controladores y aumenta la sensación de inmersión y realismo.
- Aíslan los sonidos del mundo real y esto, combinado con el uso de audio espacial, proporciona una sensación acústica muy semejante a experiencias reales.

Cuando hablamos de entornos “multimodales” nos referimos a entornos de RV que no solo ofrecen estímulos visuales o acústicos, sino que permiten simular otros estímulos como pueden ser: el tacto, olores, temperatura u otros parámetros ambientales, el gusto, la sensación de movimiento, etc.

En el contexto de la RV, también es común el uso de hardware adicional para incrementar la sensación de inmersión y realismo, ofreciendo otros estímulos e interacciones multimodales:

- **Guantes hápticos:** Son guantes con servomotores que permiten ejercer presión controlada en las manos. Esto permite simular la sensación de agarrar objetos físicos en entornos virtuales. Por ejemplo, si el usuario agarra una esfera en el entorno simulado, los guantes impedirían que cerrase del todo la mano, proporcionando una fuerza de retroalimentación. Podemos ver un guante de este tipo en la Ilustración 2.2 - Guantes Hápticos.
- **Cintas de correr omnidireccionales:** son plataformas sobre las que el usuario puede realizar las acciones de andar o correr, sin desplazarse del sitio donde se encuentra. Permiten al usuario sentir que se puede desplazar libremente dentro de la simulación sin necesidad de usar mandos para controlar el movimiento. La Ilustración 2.3 muestra algunos de estos dispositivos.

- **Otros:** hay más dispositivos pensados para mejorar la sensación de inmersión, como emisores de olores, simuladores de clima, del gusto, de la temperatura, del movimiento de vehículos, etc. No obstante, su uso está menos extendido. En este trabajo se intentó utilizar un emisor de aromas de la empresa olorama [28].



Ilustración 2.1 - Ejemplo de HMD y Controladores en entornos de RV



Ilustración 2.2 - Guantes Hápticos



Ilustración 2.3 - Cintas Omnidireccionales para caminar en entornos de RV

2.2. Computación afectiva

La computación afectiva es un campo interdisciplinar que comprende el área de la informática (con un gran peso en el subcampo de la Inteligencia Artificial), la psicología y la ciencia cognitiva, entre otros, y que consiste el estudio y desarrollo de sistemas capaces de reconocer, interpretar, procesar y hasta simular emociones humanas y empatía [65] [16] [20] [18].

2.2.1. Aplicaciones actuales

Detección y reconocimiento de emociones

Consiste en la detección de información emocional en usuarios, es decir, reconocer la emoción que están experimentando, como por ejemplo: miedo, ira, tranquilidad, calma, estrés, etc. Usualmente se utilizan sensores pasivos que recopilan datos de todo tipo sobre el estado físico del usuario y su comportamiento, que finalmente son interpretados usando distintos sistemas y modelos.

Los datos recopilados para este cometido son muy diversos y de distintas clases, incluyendo:

- Grabación en video con capturas de los gestos faciales del usuario, su postura corporal y sus gestos. [65] [32] [39] [82]
- Grabación de audio con la voz del usuario y fragmentos de conversaciones suyas para estudiar su forma de hablar, características únicas, etc. [13]
- Análisis de texto escrito (por ejemplo, mensajes en redes sociales) para inferir la emoción experimentada. [5]

- Datos fisiológicos de distinta naturaleza como la respuesta galvánica o eléctrica de la piel (GSR), la temperatura de la piel, la presión arterial, el ritmo cardíaco (HR), etc. [60] [4] [12] [14] [17] [27]

Una vez recopilada la información necesaria, es necesario utilizar distintas técnicas y algoritmos para poder clasificar las distintas emociones. Normalmente se usan técnicas de Machine Learning o aprendizaje no supervisado.

Algunas técnicas que se puede calificar como de computación afectiva son:

- Reconocimiento de voz
- Procesado natural del lenguaje
- Detección de expresiones faciales
- Sistemas de clasificación emocional a partir de datos fisiológicos.

El objetivo final es la clasificación fiable y precisa de la emoción experimentada por un usuario en un momento, o ante un estímulo, concreto.

Simulación de Emociones

Otro objetivo de la computación afectiva es la creación de sistemas capaces de simular emociones humanas creíbles, con el objetivo de mejorar la comunicación e interacción hombre-máquina. Algunos ejemplos son:

- Simulación de expresiones faciales: aplicado a robots y en interfaces con pantalla. [71]
- Simulación de reverberaciones de voz: aplicado ampliamente por compañías telefónicas para la creación de locutores automáticos que dan la sensación de ser humanos. [31]
- Simulación de texto escrito por humanos: utilizado en noticias creadas automáticamente sin la intervención de seres humanos, sino a partir de información (por ejemplo, noticias financieras), o la creación de bots. [69]

2.2.2. Aplicaciones Futuras Posibles

Si bien la computación afectiva tiene muchas aplicaciones actuales y está extendida en diversos ámbitos, todavía tiene mucho potencial en cuanto a desarrollo y a su explotación. A continuación, se enumeran algunos ámbitos en los que la computación afectiva podría aportar beneficios, destacando algunos casos de uso concretos de aplicación. Si bien es cierto que estos ámbitos ya se han tratado de explorar, todavía queda camino por recorrer para su la implantación efectiva de la computación afectiva en los mismos. [48] [21]

Educación

Una de las aplicaciones más interesantes está en el ámbito de la educación. La computación afectiva se puede utilizar para adaptar la docencia al ritmo de aprendizaje y las características personales de cada individuo. Por ejemplo, si se detecta que un método de enseñanza provoca una respuesta afectiva negativa en el alumno (empeoramiento del ritmo de aprendizaje, estrés, bloqueo, etc.), se puede introducir un cambio de forma temprana y mejorar la experiencia del individuo.

En un futuro, con el incremento de docencia online o no supervisada, esta aplicación de la computación afectiva puede ser muy importante. Los programas de educación a distancia pueden ofrecer al alumno programas distintos de educación (aunque todos cumpliendo los mismos requisitos) que se adapten a la forma de aprender. [77]

Dentro de este campo, también se puede usar para trabajar con personas con **autismo** para desarrollar tratamientos y estudios que finalmente se traduzcan en una mejora de su calidad de vida. Un ejemplo de aplicación son los videojuegos con soporte emocional [34]

Seguridad

Dentro del ámbito de la seguridad, la computación afectiva tiene varias aplicaciones, del que se destacan dos ejemplos concretos.

Por un lado, en la seguridad durante la conducción, vehículos equipados con sistemas de reconocimiento emocional podrían ser capaces de detectar emociones extremas como ira o estrés que pongan en riesgo la seguridad del conductor y el resto de los usuarios. De esta forma se podrían tomar medidas para evitar accidentes. [12]

Por otro lado, puede servir para la detección de depresión continuada en personas para poder evitar el deterioro de la salud mental y tragedias. Ello también aplica para el control del estado de salud en general. [21]

Ámbito privado

En la empresa privada se puede aplicar la computación afectiva para distintos fines. Con el permiso del usuario, se pueden recopilar datos para la elaboración de sistemas de recomendación de productos, o para estudiar el grado de satisfacción con los mismos. Plataformas de oferta de material audiovisual (Netflix, HBO, etc.) pueden aplicarla para estudiar la respuesta de los usuarios a distintas películas y estudiar gustos de los usuarios y mejorar sus sistemas de recomendación.

El estudio de los sentimientos de los usuarios también puede servir a empresas privadas para conocer el grado de satisfacción de sus empleados, o se puede utilizar para evitar fraudes en el caso de compañías aseguradoras.

Malos usos

La computación afectiva también se puede aplicar para fines de dudosa moralidad. Un ejemplo es el de creación de perfiles por parte de los gobiernos para estudiar a sus ciudadanos, como por ejemplo para la predicción de intención de voto, o el intento de encontrar gente opuesta a determinadas medidas.

Las empresas privadas también son candidatas para hacer mal uso de esta tecnología para la elaboración de perfiles sin el conocimiento de los usuarios, la recopilación de datos de estos, y su posterior venta a interesados. Estos datos podrían usarse para encontrar personas vulnerables, por ejemplo, en el caso de estafadores o sectas.

2.3. Representaciones de la respuesta emotiva

Existen distintas formas de representar emociones mediante distintos espacios y dimensiones. A continuación, explicamos los modelos más utilizados en la literatura.

2.3.1. Circumplex model

El modelo circumplejo fue desarrollado por James Russell en 1980 [55] y es uno de los más usados debido a su simplicidad. Representa las emociones en un espacio circular bidimensional. Sus dimensiones son Arousal y Valence:

- Valence: el eje horizontal. Representa como de “positiva” es la emoción. En el lado negativo encontramos emociones como miedo o rabia, y en el positivo encontramos emociones como felicidad.
- Arousal: el eje vertical. Representa el estado de “excitación”. En el polo positivo encontramos emociones como la sorpresa, y en el polo negativo emociones como la calma.

La combinación de estas dos dimensiones permite la representación diferenciada de gran cantidad de emociones distintas. En la Ilustración 2.5 [16] y la Ilustración 2.4 [39] podemos ver representaciones gráficas de este modelo, donde se representan las emociones principales, que se dividen en cuadrantes atendiendo a valores positivos o negativos de Arousal y Valence. Los cuadrantes serán denominados en este documento como: HVHA, HVLA, LVLA, y LVHA, atendiendo a si la Valencia (V) y el Arousal (A) son positivos (High), o negativos (Low).

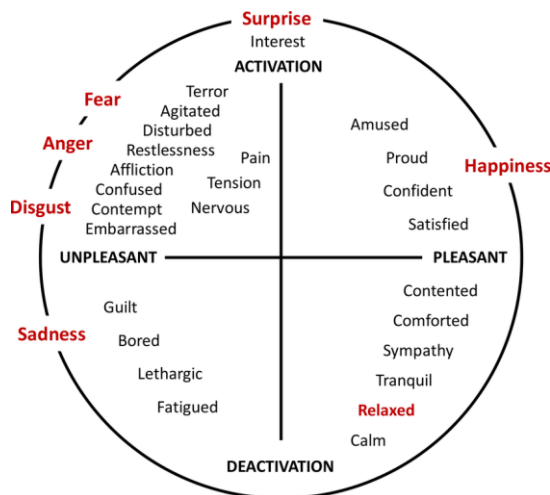


Ilustración 2.4 - Modelo Circunflejo de Russell con emociones principales

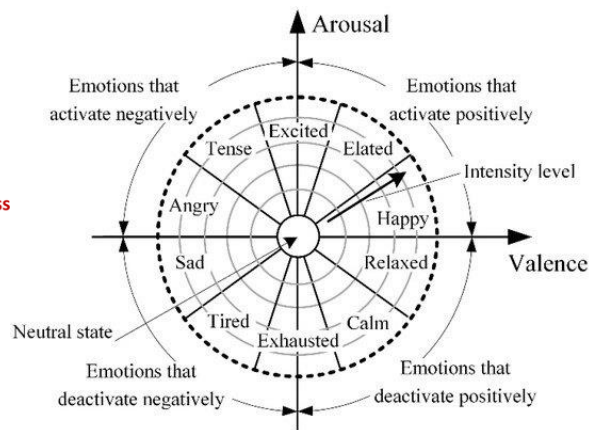


Ilustración 2.5 - Modelo Circunflejo de Russell por regiones

2.3.2. Vector Model

Vector Model es un modelo bidireccional compuesto por vectores que apuntan en 2 direcciones y que toma una forma de “boomerang” (Ver Ilustración 2.6) [54]. El modelo asume que siempre hay una dimensión de excitación subyacente (Arousal) y la dimensión Valence que determina la dirección. De esta forma, emociones con un alto grado de excitación están bien diferenciadas según son positivas o negativas (Valence), pero aquellas más calmadas no son tan diferenciadas, como por ejemplo calma y tristeza, que estarían muy próximas en este modelo.

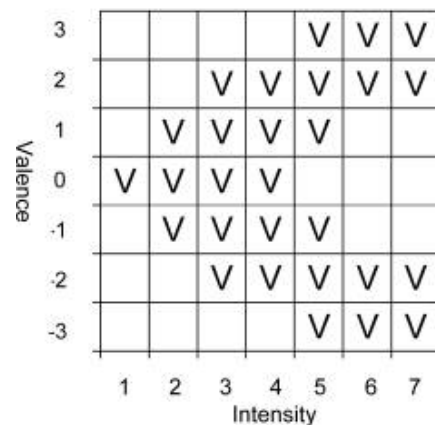


Ilustración 2.6 - Modelo Bidireccional Vector Model

2.3.3. Pana model

El modelo PANA (Positive Activation – Negative Activation) fue creado en 1985 por Watson y Tellegen [72]. Se basa en 2 direcciones: efecto positivo, y efecto negativo (Ilustración 2.7 [58]).

Es semejante al modelo circunplejo pero con los ejes girados 45º (Ilustración 2.8). Al igual que en el modelo de vectores, las emociones con un alto grado de excitación están más diferenciadas que aquellas más calmadas.

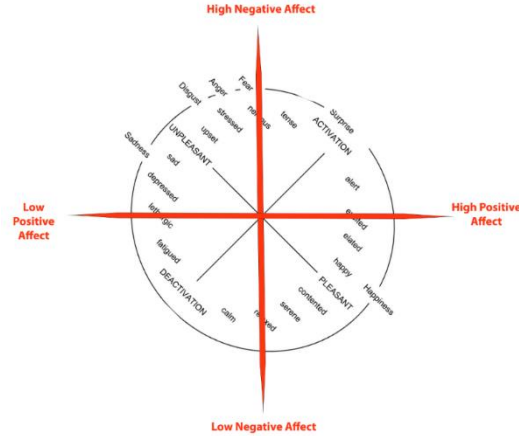
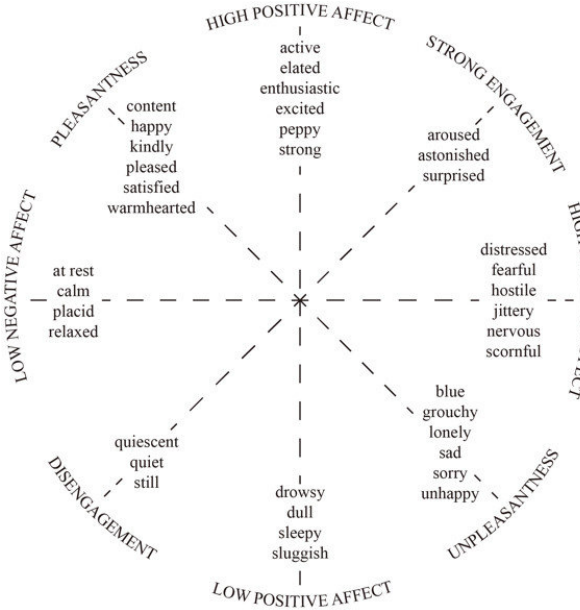


Ilustración 2.7 - Modelo PANA

Ilustración 2.8 - Modelo PANA representando como modelo circunplejo girado 45º

2.3.4. Plutchik's model

El modelo de Plutchik es un modelo de representación de las emociones tridimensional. Las emociones están organizadas en círculos concéntricos, siendo más básicas las emociones que se encuentran en los círculos internos y más complejas aquellas en los externos (Ilustración 2.10) [5]. Combinaciones de emociones diadas (“dyads” originalmente en inglés) [37] [61] son posibles mezclando emociones distintas como podemos ver en la Ilustración 2.9 [75].

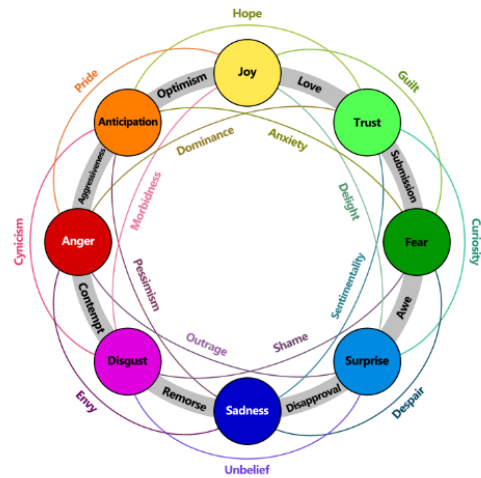
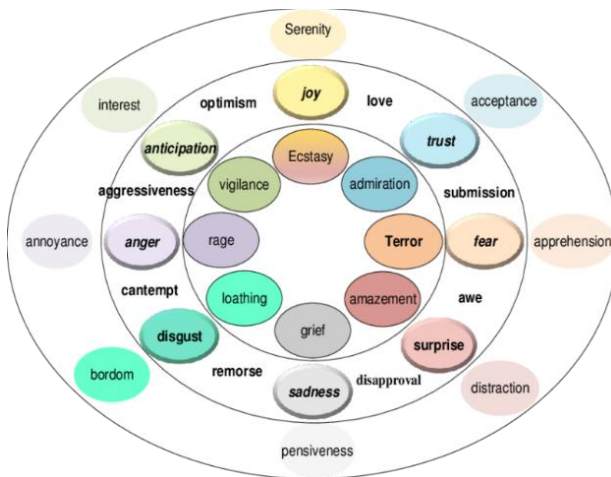


Ilustración 2.10 - Modelo de Plutchik

Ilustración 2.9 - Diadas del modelo de Plutchik

2.3.5. PAD Model

Creado por el mismo autor que el modelo circunplejo, el PAD Model representa las emociones de forma tridimensional mediante las dimensiones: Pleasure, Arousal, y Dominance, de ahí su acrónimo PAD [37]. Es semejante al modelo circunplejo, pero con una dimensión extra: Dominance, que representa la naturaleza dominante de la emoción. Por ejemplo, la rabia es una emoción dominante, mientras que el miedo no lo es, siendo ambas emociones negativas (Pleasure Negativo). En las ilustraciones Ilustración 2.11 [82] e Ilustración 2.12 [18] podemos ver un mapeado de las distintas emociones mediante el modelo PAD.

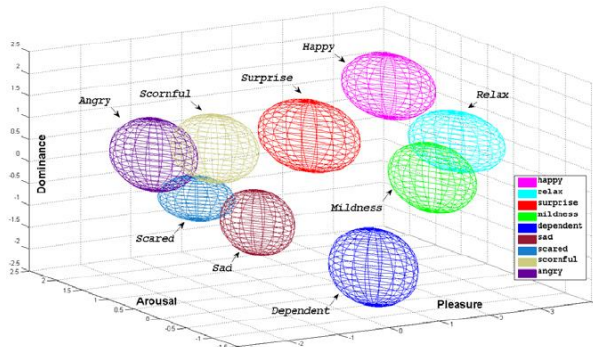


Ilustración 2.11 - Áreas que ocupan distintas emociones en modelo PAD

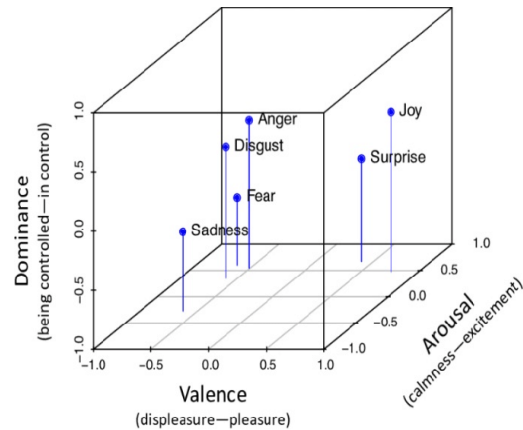


Ilustración 2.12 - Puntos principales de distintas emociones en modelo PAD

2.4. Estudios en el campo de la computación afectiva relacionados

Uno de los objetivos propuestos durante la elaboración del presente trabajo fue el de aumentar el dataset que creásemos con datos de datasets de estudios similares en el campo de la computación afectiva. A continuación, repasamos algunos de los estudios en este ámbito consultados.

2.4.1. AffectiveROAD

Se trata de un estudio dedicado a analizar el estrés en conductores, en condiciones reales. Se utilizaron sensores no invasivos para recopilar datos fisiológicos de los voluntarios mientras se les pedía que realizasen distintos trayectos. Cada viaje se realizaba en distintas condiciones de velocidad, condiciones ambientales del interior del vehículo, y del trayecto [24] [25].

No utilizamos el dataset de dicho estudio porque, aunque sea interesante, las métricas tomadas (estrés) no son del todo compatibles con nuestro objeto de estudio (emociones).

2.4.2. Amigos

Se trata de un estudio de afecto, rasgos de la personalidad y humor utilizando distintas señales neuro-fisiológicas. Como rasgo diferenciador de este estudio, los voluntarios vieron videos tanto cortos como largos, de forma tanto individual como en grupo. El objeto de este estudio era analizar las respuestas emotivas de los voluntarios en relación con su personalidad y humor en presencia de otros voluntarios [43].

Las anotaciones de las emociones se hicieron tanto de forma auto-evaluada por los voluntarios, como anotada por los compañeros de los voluntarios. Se recogieron gran variedad de métricas como Arousal, Valence, Control, familiaridad, etiquetas de emociones básicas, etc.

No obstante, no hemos podido usar su dataset porque es privado, y tras solicitar acceso a los datos tras firmar un EULA (End-User License Agreement), nos denegaron el acceso.

2.4.3. CASE

Este estudio utiliza videos para incitar emociones en los sujetos, tomando anotaciones contiguas usando el modelo circuplejo (Arousal/Valence).

El dataset CASE (Continuously Annotated Signals of Emotion) se confeccionó con el objetivo principal de ser útil para el estudio de las emociones. Consiste en señales fisiológicas de diversa índole tomadas en tiempo real y anotadas en función de Arousal y Valence, por lo que es compatible con el modelo circuplejo de Russell [59].

Lamentablemente, y a pesar de ser un dataset público, no pudimos utilizarlo por incompatibilidad a la hora de tomar las medidas de datos como Galvanic Skin Response (GSR). Ello es debido a que el método utilizado en dicho dataset consiste en la diferencia entre los valores tomados por sensores colocados en el pie y en la mano, mientras que el sensor GSR del que disponíamos solo permitía medir datos entre dedos de la misma mano o entre ambas manos.

2.4.4. Cognitive Load Inference

Consiste en un dataset confeccionado para estudiar diversos aspectos como la carga cognitiva al realizar distintas tareas y rasgos de la personalidad de los participantes en distintas dimensiones:

- **Medidas de honestidad-humildad:** sinceridad, equidad, evitación de la codicia, modestia.
- **Medidas de emocionalidad:** miedo, ansiedad, dependencia, sentimentalismo.
- **Medidas de extraversión:** autoestima social, audacia social, sociabilidad, vivacidad.
- **Medidas de amabilidad:** perdón, mansedumbre, flexibilidad, paciencia.
- **Medidas de conciencia:** organización, diligencia, perfeccionismo, prudencia.
- **Medidas de apertura a la experiencia:** estética, curiosidad, creatividad, poco convencionalismo.

Si bien es un estudio muy completo y de libre acceso, las métricas emocionales recopiladas no son compatibles con nuestro objetivo [22].

2.4.5. DEAP

DEAP (Database for Emotion Analysis using Physiological Signals) es un dataset donde los participantes fueron expuestos a videos musicales de una emisora de radio, que calificaron usando las escalas de Arousal y Valence a través de etiquetas emocionales [36] [8].

Utiliza gran cantidad de señales fisiológicas compatibles con nuestros sensores, usa el modelo circuplejo, y una metodología en la confección del dataset semejante a la nuestra. No obstante, se trata de un dataset privado al que no nos concedieron acceso y no pudimos utilizar.

2.4.6. Dreamer

Para la confección de este dataset, los participantes fueron sometidos a estímulos audiovisuales y se midieron sus emociones según Valence, Arousal y Dominance (Circumplex model + dominance). No obstante, se utilizaron datos fisiológicos para cuya adquisición no contábamos con los sensores adecuados, además de no recopilar otras métricas como GSR. Por tanto, no fue posible usarlo [33].

2.4.7. MAHNOB

Consiste en un dataset semejante al de Dreamer y DEAP: se hizo con participantes viendo videos y anotando sus emociones en las dimensiones de Arousal y Valence. Los participantes eran además monitorizados con videocámaras, micrófonos y rastreadores de mirada (eye gaze tracker), aparte de recopilar gran cantidad de datos fisiológicos (electrocardiograma, encefalografía de 32 canales, amplitud de la respiración y temperatura de la piel) [63].

El dataset sin embargo es privado, y al igual que con otros, no nos concedieron acceso al mismo, por lo que no pudimos usarlo.

2.4.8. Resumen de datasets de estudios similares

Resumen de datasets consultados

Datasets para tener más datos y crear modelos.

Título/Nombre	Tipo Acceso	Representación	Acceso Obtenido	Usado
Affective Road	Libre	Otra: Estrés	---	No
AMIGOS	Privado	Circumplex	No	No
CASE	Libre	Circumplex	---	No
Cognitive Load	Libre	Otra: carga cognitiva	---	No
DEAP	Privado	Circumplex	No	No
Dreamer	Privado	Circumplex	Si	No
MAHNOB	Privado	Circumplex	No	No

Tabla 2.1 - Datasets de Estudios Similares

Datos recopilados en datasets a los que hemos tenido acceso:

Nombre	HR/ECG	TEMP	EDA/GSR	Otros	Sujetos/Tests
Affective Road	Sí	Sí	Sí / No	Acc, BVP, BR	13
CASE	Sí	Sí	Sí	BVP, Resp	30
Dreamer	Sí	No	No	EEG	23

Tabla 2.2 - Métricas Clave de datasets Existentes

2.5. Estudios relacionados que integren computación afectiva con entornos de RV multimodales.

En esta parte consultaremos estudios que combinan computación afectiva con entornos de realidad virtual para distintos fines.

2.5.1. Estudio de la respuesta emotiva en escenarios de RV

En esta parte nos apoyamos en el estudio realizado en septiembre de 2020 por Marín-Morales et al. [41]. En dicho estudio se realiza un análisis de los distintos artículos publicados hasta la fecha que tratan sobre computación afectiva utilizando herramientas de RV. De 1425 artículos consultados en el estudio, únicamente 42 pasaron el filtro utilizado para catalogarlos como relacionados con Computación afectiva y RV. Se filtró aquellos artículos que no utilizaban HMD y entornos de RV, los que no tenían como estudio las emociones o similar, y los que no utilizaban ninguna métrica fisiológica o biométrica.

Los primeros artículos que mezclan computación afectiva con entornos de RV datan de 2002. El más antiguo, de Jang et al. [30], estudia la respuesta fisiológica en entornos de RV donde se simulan tareas de conducción y aviación y se recogen distintas métricas como GSR o el HR, y se estudian los efectos producidos en los usuarios como mareos, o el grado de excitación (Arousal).

En los primeros artículos de este campo se intentan analizar cómo influyen el uso de escenarios inmersivos de RV emocionalmente en los usuarios en forma de excitación y “grado de presencia” [53] [76].

Estudios posteriores, como por ejemplo Moghimi et al. (2016) [44], empiezan a analizar la respuesta emotiva de los usuarios a distintos eventos en entornos de RV utilizando distintas dimensiones para clasificar las emociones como Arousal, Valence y Dominance, siendo el uso de cuestionarios a usuarios la herramienta más frecuente para la clasificación de la emoción.

En estudios más recientes como Marín-Morales et al. (2018) [40], se utilizan sensores “wearables”, es decir, dispositivos capaces de recopilar métricas, pero a la vez fáciles de portar, utilizar, con un menor consumo y más asequibles que los sensores tradicionales más restrictivos. En Toet et al. (2020) [67], se propone una alternativa a cuestionarios para evaluar la emoción experimentada por los usuarios: una forma de evaluación continua utilizando una herramienta nueva llamada EmojiGrid, donde el usuario en el modelo de representación de las emociones circunflejo de Russell señala cómo se siente en cada momento.

En general, en estudios recientes se suelen utilizar para tareas predicción de las emociones herramientas de inteligencia artificial y aprendizaje máquina de clasificación y regresión como: SVM, Random Forest, modelos de redes neuronales, modelos de RNN o modelos de LSTM. [4] [40] [79]

Un problema que encontramos al analizar los distintos artículos es que apenas hay datasets disponibles abiertos. De los 42 artículos analizados en Marín-Morales et al. [41], solo 4 estudios hacen públicos los datasets creados.

2.5.2. Uso de computación afectiva combinada con escenarios de RV como herramienta de apoyo al aprendizaje

Durante la introducción a la computación afectiva realizada previamente en este documento, listábamos una serie de posibles aplicaciones futuras de la computación afectiva entre las que destacábamos su potencial como herramienta de apoyo al aprendizaje.

Por ejemplo, en Ontiveros-Hernández et al. [48] se realiza un estudio donde se utilizan escenarios de RV para tareas de aprendizaje como entrenamiento en trabajos de mantenimiento de líneas de alta tensión de forma segura en entornos virtuales simulados. Las emociones de los usuarios en cada tarea y sus tasas de aprendizaje fueron analizadas para estudiar a qué eventos respondían mejor los usuarios y adaptar la forma de entrenamiento.

2.5.3. Uso de computación afectiva combinada con escenarios de RV como herramienta de rehabilitación

Previamente comentamos el potencial de la computación afectiva combinada con el uso de entornos de RV en el campo de la medicina o la psiquiatría.

En Girishan Prabhu et al. [21], se trata el problema de la ansiedad y el dolor experimentado por pacientes sometidos a cirugías antes y después de las intervenciones. Se propone el uso de entornos de RV como alternativa al uso de opioides para tratar la ansiedad y el dolor, pues estos provocan adicción en 1 de cada 16 pacientes. Se propone además un sistema de computación afectiva que monitorice los datos fisiológicos del paciente para adaptar la experiencia virtual a las necesidades del paciente en cada momento.

2.6. Modelo LSTM de red neuronal

En tareas de machine learning consistentes en clasificación multiclase o regresión el uso de redes neuronales como herramienta está muy extendido. En este trabajo hemos utilizado una red neuronal con arquitectura LSTM como herramienta de regresión para estimar las emociones experimentadas por los usuarios en base a los datos fisiológicos extraídos.

En esta parte explicaremos de forma breve en qué consisten las redes neuronales, las redes neuronales recurrentes (a partir de ahora RNN), y la arquitectura LSTM.

2.6.1. Redes neuronales

A continuación, explicamos el concepto de redes neuronales de forma breve.

Las redes neuronales son un modelo computacional cuyos elementos principales son:

- Neuronas artificiales: una red neuronal es un grafo de neuronas artificiales, que son unidades que reciben una serie de entradas (inputs) sometidas a una serie de pesos. Aplican una función de activación a las entradas recibidas y devuelven un valor de salida (output). Ilustración 2.13 [81]

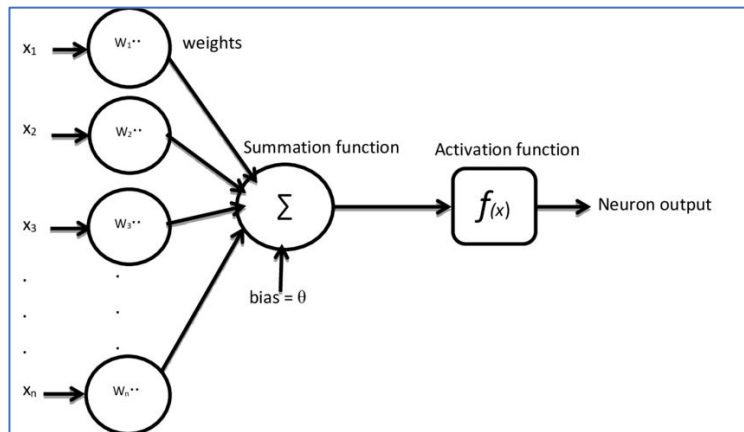


Ilustración 2.13 - Estructura de neurona artificial simple

- Conexiones y pesos: en una red neuronal, las neuronas están conectadas unas a otras siguiendo diversas topologías (Ver Ilustración 2.14 [74]). Cada conexión transfiere la salida de una neurona a la entrada de otra con un peso asociado que puede variar.

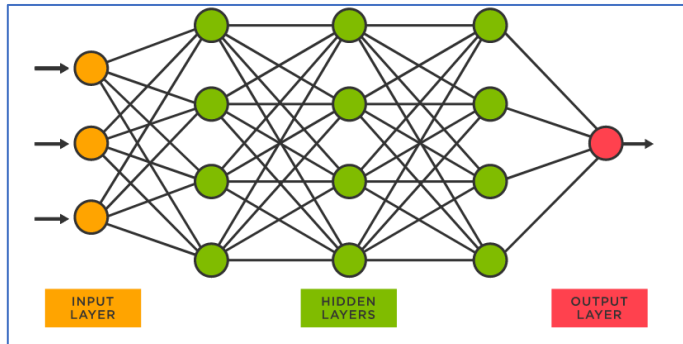


Ilustración 2.14 - Estructura típica de red neuronal

- Función de activación: es la función utilizada por las neuronas para procesar las entradas recibidas y generar una salida. Algunas de las funciones de activación más usadas son la activación sigmoide, lineal, ReLU, softmax, o la tangente hiperbólica.

- Función de aprendizaje y error: a partir del error calculado a la salida de la red neuronal, la función de aprendizaje es la encargada de retro propagar el error y modificar los pesos de las conexiones entre neuronas para conseguir que la red “aprenda” y mejore la salida en cada época de entrenamiento.

Las redes neuronales normalmente están estructuradas por capas, con una capa de entrada, varias capas ocultas y una capa final de salida. Las conexiones entre capas varían según el tipo de arquitectura usada.

2.6.2. Redes neuronales recurrentes

Las RNN son un tipo de arquitectura donde la salida de una neurona es reutilizada como entrada en una neurona de una capa anterior o sobre la misma neurona.

Esto permite a las redes neuronales trabajar con datos secuenciales donde es necesario tener en cuenta datos anteriores y no únicamente la última entrada a la red (ver Ilustración 2.15 [52]).

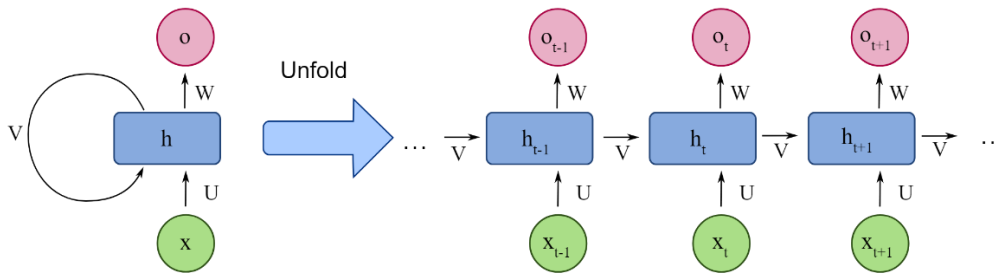


Ilustración 2.15 - Red Neuronal Recurrente

2.6.3. Modelo LSTM

Una red Long Short-Term Memory (LSTM) es un tipo de RNN especial que trabaja muy bien con problemas de predicción de secuencias. En nuestro caso trabajamos con datos secuenciales, en orden cronológico y que dependen en gran medida de estados anteriores, por lo que el modelo LSTM parece el idóneo para nuestro problema.

Diferencias respecto a redes recurrentes normales

En los modelos de RNN la información puede permanecer en la red gracias a que existen bucles entre capas ocultas que conectan capas con capas anteriores y por tanto permiten recordar estados previos. Esto se consigue en las redes recurrentes normales modificando las conexiones entre neuronas y capas.

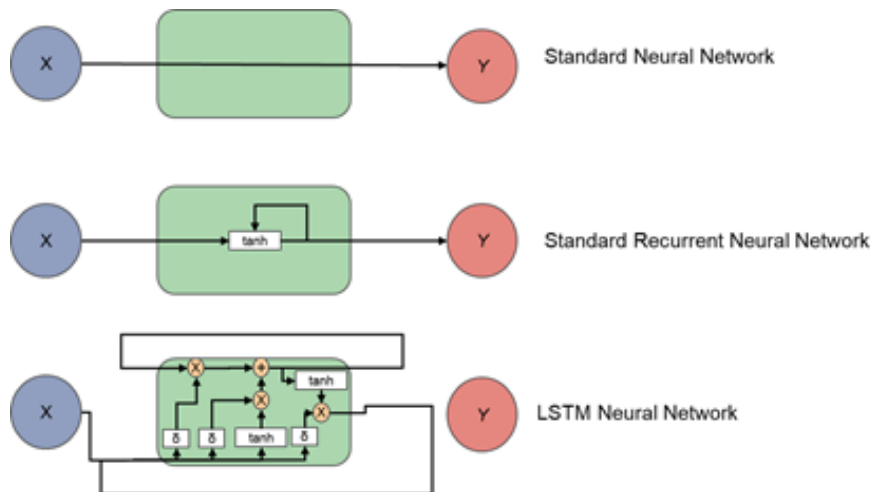


Ilustración 2.16 - Comparación simplificada de distintas arquitecturas de redes neuronales

Un problema de las RNN, es que los datos pasados tienen menos peso que los datos más recientes debido al problema del desvanecimiento del gradiente. Como consecuencia, puede ser complicado entrenar RNN para resolver problemas con dependencias temporales extensas.

En el caso de las redes LSTM contamos con una “celda de memoria” que pueden mantener información a largo plazo, lo que permite trabajar con secuencias temporales largas sin pérdida de peso en datos pasados. Esta “celda de memoria” consiste en un conjunto de “puertas” (gates) cuya función es decidir qué información ingresa a la memoria, qué información sale, y qué información es olvidada (ver Ilustración 2.16 [1] e Ilustración 2.17 [73]).

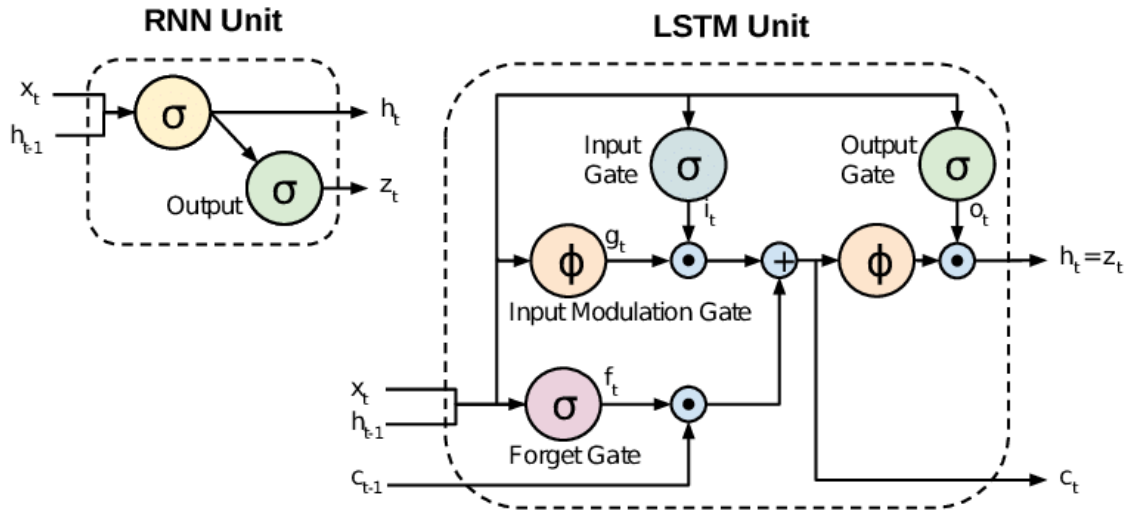


Ilustración 2.17 - Diferencias RNN y LSTM

Funcionamiento

A continuación, explicamos en detalle el funcionamiento de cada unidad LSTM [9].

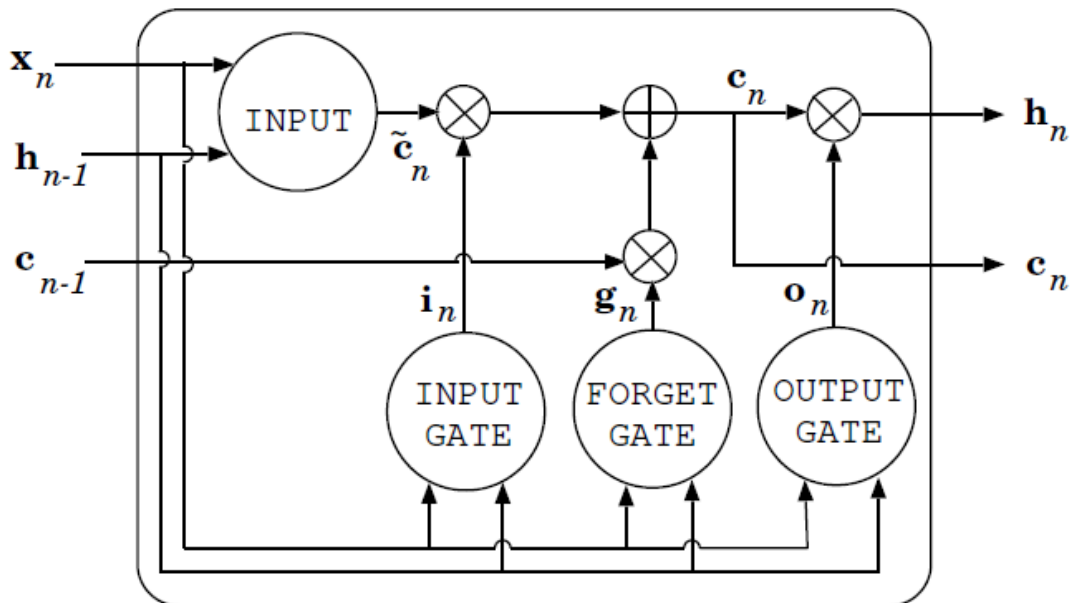


Ilustración 2.18 - Esquema LSTM

En primer lugar, la salida \tilde{c}_n es el **equivalente a la salida de** una neurona en una **RNN** normal. Toma como entradas los datos nuevos, y la hipótesis del estado anterior, aplica una matriz de

pesos a ambos, y devuelve una salida. En este caso utilizamos la función de activación tangente hiperbólica (Tanh) porque permite valores negativos y positivos, y ayuda a evitar problemas de desvanecimiento del gradiente. \tilde{c}_n se calcula aplicando la Ecuación 2.1. .

$$\tilde{c}_n = \mathbf{f}_{th}(\mathbf{W}_H^C \mathbf{h}_{n-1} + \mathbf{W}_X^C \mathbf{x}_n)$$

Ecuación 2.1 - Ecuación Funcionamiento LSTM 1

A continuación, pasamos un primer **filtro en la puerta de entrada** (input gate) donde se decide a partir de las mismas entradas y estado anterior si aceptamos la entrada. A la entrada y a la hipótesis anterior se le aplica una matriz de pesos y con una función de activación sigmoide. La función de activación sigmoide tiende a cero cuando encuentra valores negativos.

El filtro consiste en realizar un producto elemento a elemento (element-wise product) entre la salida \tilde{c}_n calculada anteriormente y salida \dot{i}_n de la puerta de entrada. Si el valor de \dot{i}_n es nulo, el resultado de esta operación será nulo y por tanto estamos evitando que la entrada continúe su proceso. \dot{i}_n se calcula aplicando la Ecuación 2.2 .

$$\dot{i}_n = \mathbf{f}_s(\mathbf{W}_H^I \mathbf{h}_{n-1} + \mathbf{W}_X^I \mathbf{x}_n)$$

Ecuación 2.2 - Ecuación Funcionamiento LSTM 2

A continuación, pasamos al **filtro realizado por la puerta de "olvido"** (forget gate). En esta puerta se decide si dejar pasar o no la información del estado interno de la memoria en el instante anterior c_{n-1} . Calculamos g_n a partir de la entrada actual y de la hipótesis del estado anterior con su matriz de pesos, y aplicamos una función de activación sigmoide. Si el resultado de la función de activación es cero, dado que tenemos un producto elemento a elemento, desechamos la información del estado interno anterior. Después tengamos un valor nulo o no, hacemos una suma elemento a elemento con lo que haya dejado pasar la puerta de entrada (que puede ser nulo o no). De esta forma controlamos si utilizamos la información del estado interno anterior o no. g_n y c_n se calculan aplicando la Ecuación 2.3 y la Ecuación 2.4 .

$$g_n = \mathbf{f}_s(\mathbf{W}_H^F \mathbf{h}_{n-1} + \mathbf{W}_X^F \mathbf{x}_n)$$

Ecuación 2.3 - Ecuación Funcionamiento LSTM 3

$$c_n = g_n \times c_{n-1} + \dot{i}_n \times \tilde{c}_n$$

Ecuación 2.4 - Ecuación Funcionamiento LSTM 4

Finalmente pasamos **el filtro de la puerta de salida** (output gate). El funcionamiento es igual que en el caso de la puerta de entrada, la diferencia es que aquí lo que controlamos es si dejamos pasar el estado interno actual c_n como hipótesis o no. Calculamos o_n a partir de los datos de

entrada y la hipótesis anterior, aplicamos activación sigmoide, y finalmente realizamos un producto elemento a elemento con c_n para calcular la hipótesis de salida de la unidad LSTM h_n . o_n y h_n se calculan aplicando la Ecuación 2.5 y la Ecuación 2.6 .

$$\mathbf{o}_n = \mathbf{f}_s(\mathbf{W}_Y^O \mathbf{h}_{n-1} + \mathbf{W}_X^O \mathbf{x}_n)$$

Ecuación 2.5 - Ecuación Funcionamiento LSTM 5

$$\mathbf{h}_n = \mathbf{o}_n \times \mathbf{f}_{th}(\mathbf{c}_n) = \mathbf{F}(\mathbf{x}_n, \mathbf{h}_{n-1})$$

Ecuación 2.6 - Ecuación Funcionamiento LSTM 6

3. Experimentación

En esta sección explicaremos el proceso que hemos seguido durante el proceso de creación de los datasets, la creación del modelo, el entrenamiento del mismo, y su aplicación. Veremos también las herramientas que hemos utilizado.

3.1. Librerías

El trabajo referente a los modelos de computación afectiva ha sido realizado mediante el uso del lenguaje de programación Python. A continuación, nombramos algunas de las librerías que hemos utilizado.

3.1.1. TensorFlow y Keras

Las librerías de TensorFlow [2] y Keras [11], son librerías especializadas para trabajar con problemas de Machine Learning e Inteligencia Artificial. Proporcionan una gran cantidad de métodos para la creación rápida de distintos modelos de redes neuronales con distintas arquitecturas, funciones, etc.

En nuestro trabajo hemos utilizado estas librerías para la creación de un modelo de regresión basado en una red neuronal del tipo Long Short-Term Memory (LSTM), que posteriormente entrenamos para predecir valores de Valence, Arousal, y Dominance a partir de datos fisiológicos.

3.1.2. Pandas y Numpy

Pandas [42] y Numpy [26] son librerías que sirven para trabajar con datos numéricos, matrices, etc. Facilitan además las tareas de leer archivos de datos como aquellos en formato .csv.

Las hemos utilizado para trabajar con los datos adquiridos en las pruebas.

3.1.3. Scipy y Sklearn

Al igual que Keras o Tensorflow, Scipy [70] y Sklearn [49] permiten la creación de modelos de Machine Learning. No están centradas en redes neuronales, se centran más en otros modelos como máquinas de vectores de soporte (SVM), Random Forests, y otros clasificadores. Son librerías que ofrecen además diversas funcionalidades matemáticas (p.ej. cálculo de valores medios, desviaciones, módulos, distancias entre vectores, etc.), y es por este motivo por el que las hemos usado.

3.1.4. Librerías Adicionales

Además de las ya mencionadas hemos utilizado otras librerías auxiliares, como por ejemplo Matplotlib [29] para la creación de gráficas e imágenes.

3.2. Dispositivos de captura de datos fisiológicos consultados

Para la realización de este trabajo nos fue necesario adquirir sensores para la recopilación de métricas fisiológicas. En este trabajo nos enfocamos a herramientas y dispositivos escalables y utilizables durante el uso de dispositivos de RV. Por estos motivos:

- No podemos utilizar dispositivos que requieran un gran número de sensores: los dispositivos de venta al público más vendidos no suelen ser capaces de recopilar mucha información.
- No podemos utilizar sensores invasivos que normalmente requieran usar sensores en los dedos, o cabeza: hay que tener en cuenta que el usuario puede tener las manos ocupadas para la interacción con el entorno RV, y que llevará un HMD.

- No podemos usar capturas de video para reconocer gestos dado que el tendrá como mínimo la mitad de la cara tapada a causa de llevar el HMD puesto, por lo que no es viable.

Los dispositivos más comunes que atienden a nuestros criterios son los llamados *smartwatches*, dispositivos portátiles que se conectan a los teléfonos móviles y que se llevan en forma de pulsera en la muñeca.

Nos centraremos en los datos fisiológicos más comúnmente medidos por dichos dispositivos como: el HR la temperatura de la piel, o el GSR, aunque hay que tener en cuenta que no todos los dispositivos disponibles en el mercado miden la respuesta galvánica o la temperatura de la piel.

A continuación, comentamos algunos de los dispositivos investigados y que en algunos casos intentamos utilizar, y el motivo por el cual finalmente quedaron descartados, así como los sensores que acabamos utilizando.

3.2.1. Pulseras inteligentes

Nuestra primera opción fue buscar distintas pulseras inteligentes (*smartwatches*). Nuestros requisitos eran:

- Que fueran capaces de recopilar datos como el ritmo cardíaco y la conductividad térmica (GSR o EDA).
- Que fueran asequibles (disponibles y de bajo coste).
- Que permitiese guardar los datos y acceder a ellos, no solo mostrarlos. Esto sin necesidad de contratar un servicio extra.

Entre los distintos dispositivos que encontramos destacamos por ejemplo las pulseras “Algoband F8”, “Polar M400”, “Empatica E4”, “Microsoft Band 2”.

Nombre	Precio	EDA/GSR	HR	Acceso Datos Raw
Algoband F8	~ US \$28.58	No	Sí	Contratando servicio Extra
Polar M400, Polar H7	~ US \$129	No	Sí	Contratando servicio Extra
Empatica E4	~ US \$1,690.00	Sí	Sí	Contratando servicio Extra
Microsoft Band 2	DESCONTINUADO	Sí	Sí	Desconocido

Tabla 3.1 - Pulseras Smartwatch Consultadas

Ninguna cumplió nuestros requisitos. Si bien Microsoft Band 2 parecía cumplirlos, es un producto discontinuado.

3.2.2. MySignals

Consiste en un kit de sensores para la toma de métricas fisiológicas, desarrollado por la empresa Libelium (Zaragoza) [38]. Aunque tiene un precio alto (+1.100\$), tuvimos acceso a un kit prestado por la Universitat de València.

El problema que encontramos es que si bien funcionaba correctamente y mostraba los datos de forma fiable, no era posible acceder a los datos sin contratar un servicio anual con la empresa. Además, aun en caso de contratar el servicio de pago, no permite capturas de más de 30 segundos, lo cual era incompatible con nuestro objetivo de capturar datos en largos periodos de tiempo.



Ilustración 3.1 - Kit de sensores "MySignals SW" desarrollado por la empresa Libelium

El kit al que hacemos referencia es el Kit "My Signals SW" (Ilustración 3.1), otros kits de la misma empresa como "My Signals HW" permiten la toma de datos directamente.

3.2.3. Sensores seleccionados y utilizados

Finalmente, como ninguno de los productos existentes cumplía con los requisitos del proyecto, se optó por desarrollar un prototipo propio, que incluyese los sensores buscados y permitiese el acceso a los datos capturados por los mismos. Decidimos utilizar una placa Arduino Nano con sensores conectados para la toma de métricas. Los dispositivos utilizados fueron:

- Arduino Nano: placa para conectar los sensores y recopilar los datos obtenidos (Ilustración 3.2).
- Sensor GSR (desarrollado por Sichiray): sensor para medir la respuesta galvánica de la piel (GSR). Este sensor toma las medidas conectando los bornes a los dedos de las manos (Ilustración 3.4).
- Sensor de frecuencia cardíaca "max30102" (desarrollado por Maxim Integrated): sensor para medir los latidos por minuto. Utiliza un sensor óptico para ello (Ilustración 3.3).

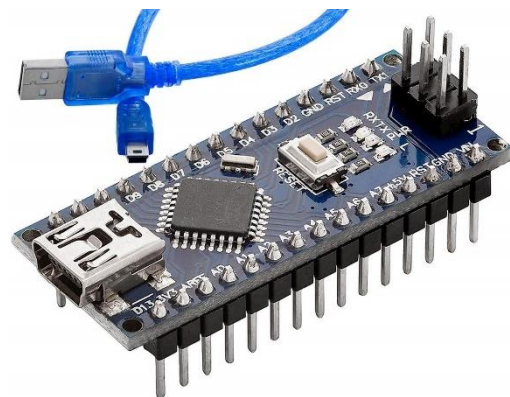


Ilustración 3.2 - Placa electrónica Arduino Nano



Ilustración 3.4 - Sensor GSR de la empresa Sichiray



Ilustración 3.3 - Sensor "Max30102" para la medición de la frecuencia cardiaca de la empresa Maxim Integrated

En el marco de este trabajo se realizaron tareas de circuitería y cableado para la creación de dicho prototipo. Esta tarea se llevó a cabo exitosamente, como demuestra el hecho de fuéramos capaces de recopilar los datos deseados.

3.3. Entornos de Realidad Virtual Creados y Utilizados

En este trabajo, hemos adoptado entornos virtuales desarrollados por la Universidad de Valencia en colaboración con la Fundación i2CAT, pero en los que el autor de este trabajo ha contribuido en todo su proceso, en el marco de dos proyectos de investigación. De hecho, se han publicado dos artículos científicos presentando dichos escenarios RV, con diferentes estrategias para la presentación de estímulos interactivos (ej. Subtitulado, puntos de interés, etc.) [47] [45].

En particular, se han adoptado dos reconstrucciones virtuales, interactivas y multimodales de dos teatros del Siglo de Oro como (ver Ilustración 3.5):

- Shakespeare Globe (Reino Unido).
- El Corral de Comedias de la Olivera (Valencia).

Características

- Son entornos completamente navegables
- Incluyen obras de teatro (actores capturados desde sala de croma, mediante vídeo inmersivos y audio espacial) que se pueden reproducir en los escenarios (ver Ilustración 3.8 e Ilustración 3.9)
- Las obras están subtituladas con diversos modos de presentación.
- Existe audio espacial
- Incluyen puntos de interés que aportan información adicional relevante sobre el teatro o entorno RV.

Estímulos distintos para estudiar

- Distintos modos de presentación de subtitulado. Ver Ilustración 3.10.
- Audio plano (Text-To-Speech) contra Audio grabado por actores.
- Puntos de interés que aportan información mediante audio o mediante texto. Ver Ilustración 3.6.

Dispositivos utilizados para la experiencia de realidad virtual

Capítulo 3 - Experimentación

- HMD Oculus Quest 2 y accesorios
- Portatil MSI compatible con aplicaciones de RV.

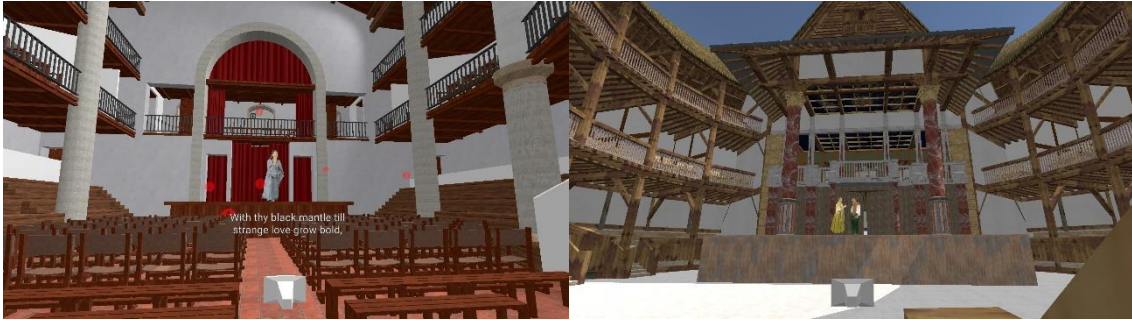


Ilustración 3.5 - Capturas de los teatros reconstruidos: (izquierda) Olivera; (derecha) Globe

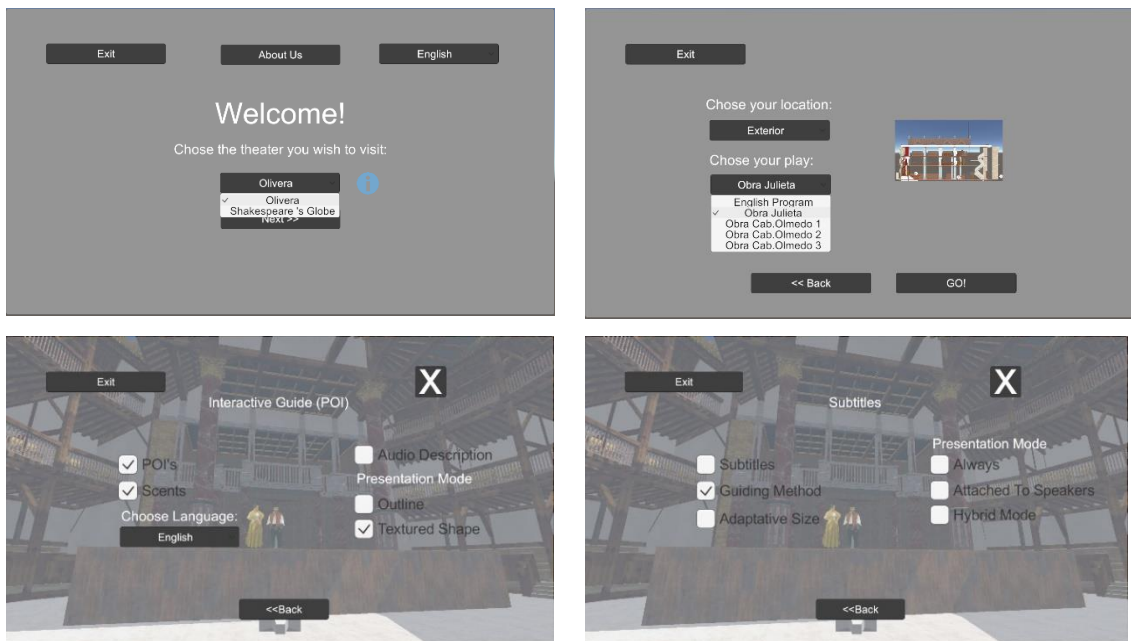


Ilustración 3.7 - Capturas de diferentes pantallas y opciones de menú



Ilustración 3.6 - Ejemplo de Punto de Interés con Texto

Capítulo 3 - Experimentación



Ilustración 3.8 - Captura de obras en sala de croma, con vídeo inmersivo 180º estereoscópico y audio espacial



Ilustración 3.9 - Vista desde balcón del escenario de RV del Corral de Comedias de la Olivera



Ilustración 3.10 - Obra con subtítulos con estilo

3.4. Creación de datasets

Para la realización de este trabajo, fue necesaria la creación de datasets que correlacionaran los datos fisiológicos recopilados con las emociones sentidas por los usuarios en cada momento. Para ello nuestra metodología consistió en pedir a los voluntarios que visualizaran una lista de videos de distintas temáticas seleccionados para evocar distintas emociones. Durante el visionado de estos tomamos métricas con los sensores. Asimismo, sometimos a los voluntarios a unos cuestionarios donde reflejaron sus emociones durante cada video.

3.4.1. Selección de videos para evocar emociones

En este trabajo decidimos utilizar el modelo de representación circunflejo de las emociones de Russell por su simplicidad, su facilidad para ser entendido, y por su capacidad de reflejar todas las emociones posibles. Además, es un modelo ampliamente utilizado en estudios similares. [40] [65] [43] [36] [33]

Recordemos que este modelo se basa en 2 dimensiones, Valence y Arousal, que reflejan la “positividad” y “excitación” de una emoción dada.

Para seleccionar una serie de videos que evocasen la práctica totalidad de las emociones, recurrimos a la literatura y seleccionamos videos atendiendo a los siguientes criterios para provocar las emociones de Ira, Asco, Sorpresa, Felicidad, Miedo, Tristeza, y Calma, que cubren todo el espacio emotivo del modelo circunflejo [62].

Ira

Para provocar esta emoción seleccionamos una serie de videos donde se aprecia violencia de distinto tipo. Por ejemplo:

- Una agresión totalmente gratuita a una pareja a manos de una serie de encapuchados. (Ilustración 3.11)
- Una escena de violación (falsa pero dura) extraída de una película.
- Fragmento de video musical animado que muestra un choque armado mientras suena música estridente y rápida.



Ilustración 3.11 - Video de agresión gratuita a pareja

Asco

Para provocar esta emoción recurrimos a videos donde se muestran mutilaciones, o invasiones de insectos. Por ejemplo:

- Un video donde se muestra como sacrifican a un grupo de cerdos infectados de forma inhumana, tirándoles a una zanja y prendiéndoles fuego todo mientras se oyen sus chillidos. (Ilustración 3.12)
- Un video de un perro en el veterinario donde se aprecia que el perro tiene multitud de gusanos enterrados bajo su piel.



Ilustración 3.12 - Sacrificio de Cerdos

Sorpresa

Para provocar esta emoción utilizamos videos donde se muestra una incongruencia o donde causamos un sobresalto en el usuario. Por ejemplo:

- Un video de animación donde no se conoce la temática, y es al final de este cuando averiguamos que se trata de un anuncio comercial.
- Un video donde provocamos un susto en el usuario. No hay que confundirlo con la emoción de miedo, un susto es una sorpresa corta.

Felicidad

Para provocar este tipo de emoción recurrimos a videos caseros donde aparecen familias interactuando de forma feliz entre ellas, videos de humor, música alegre y dibujitos animados. Por ejemplo:

- Video casero de padre jugando con su hijo
- Video humorístico
- Video de animación musical (Ilustración 3.13)
- Fragmento de la serie de dibujos “El valle de los moomins”



Ilustración 3.13 - Mongo Cat Espacial

Miedo

Para provocar esta clase de emoción intentamos que el usuario que ve el video anticipe un evento catastrófico, o que sienta una sensación de muerte súbita. Por ejemplo:

- Video de chica jugando a clavar un cuchillo entre sus dedos para causar anticipación ante posible accidente. (No ocurre)
- Video de viaje por carretera justo después de mostrar un video donde hay un accidente automovilístico acabado en tragedia.
- Videos de animación con atmósfera oscura con el cometido de causar desazón del género miedo/terror. (Ilustración 3.14)



Ilustración 3.14 - Corto animado "The Crow and the Raven"

Tristeza

Para provocar esta emoción hicimos a los usuarios visualizar videos donde aparecen personas con las que es fácil simpatizar y sentirse mal por ellas, además de videos con música triste. Por ejemplo:

- Video de entrevista a niña siria que se ha quedado huérfana en la guerra y se pone a llorar.
- Video de accidente automovilístico donde, aún sin mostrar nada, se sugiere que alguien ha muerto y la acompañante chilla de desesperación.

Calma

Para provocar esta emoción utilizamos videos con música tranquila, videos donde mostramos un proceso tranquilo, o dibujos animados. Por ejemplo:

- Video con música tranquila y pixel art.
- Video de cocina rápida donde se ven los pasos básicos para hacer varias recetas.

Capítulo 3 - Experimentación

- Dibujitos animados rusos donde suena música tranquila de acordeón.

En las pruebas utilizamos un total de 34 videos de distintas temáticas, algunos de los cuales ya hemos mencionado. La duración de los videos varía entre 5 minutos y 10 segundos, teniendo la mayoría de los videos una duración media de 30-40 segundos. Se han seleccionado con la intención de crear emociones que completen todo el espacio del modelo de representación de las emociones de Russell. En la selección de los mismos nos hemos basado en criterios de la literatura [10] [3] [35] [23] como elegir dibujos animados o videos caseros para provocar emociones positivas, videos de violencia para causar miedo o rabia, o música y escenas tristes para provocar tristeza.

Descripción del video	Duración (mm:ss)	Valencia Objetivo	Arousal Objetivo	Emoción Objetivo
Fragmento de El Valle de los Moomins (Dibujos animados)	05:52	Positiva	Negativo	Calma/Felicidad
Cheburashka - Krokodil Gena (Dibujos animados)	02:08	Positiva	Negativo	Calma
Fragmento del video musical de [METRIC - Lost Kitten]	00:43	Negativa	Negativo	Tristeza
Fragmento de canción [90sFlav - Call me] con Pixel Art	00:44	Negativa	Negativo	Tristeza
Fragmento de video casero entre abuela y nieto con atmósfera relajada.	00:37	Positiva	Negativo	Calma/Alegría
Video de accidente en carretera.	00:28	Negativa	Negativo/Neutro	Tristeza/Ansiedad
Video con musica triste	00:28	Negativa	Negativo	Tristeza
Sketch corto donde el protagonista se lleva una decepción.	00:29	Neutro	Positivo	Expectación
Grabación donde se ve una agresión gratuita a una pareja de viandantes	00:42	Negativa	Positivo	Miedo/Enojo
Video casero donde la protagonista juega a clavar un cuchillo entre sus dedos.	00:22	Neutro Negativa	Positiva	Miedo / Expectación.
Video casero donde se aprecia un perro en una consulta de veterinario. El perro tiene gusanos vivos en la piel.	00:18	Negativa	Positivo	Repulsión
Escena ficticia de la película [Irreversible (2002)] donde se escenifica una violación	00:45	Negativa	Positivo	Miedo/Enojo

Capítulo 3 - Experimentación

Fragmento de entrevista a niña siria huérfana, víctima de una guerra.	00:22	Negativa	Negativo	Tristeza
Grabación donde se aprecia un sacrificio brutal a una piara de cerdos enfermos ejecutada mediante inmolación.	01:25	Negativa	Positivo	Repulsión
Corto animado [Objectionable Apparatus 2] del artista "kol below"	04:50	Negativa	Positiva	Miedo / Desazón
Video musical de [Anibal Bley - Sanctum Regnum]	00:33	Negativa	Positiva	Miedo / Desazón
Corto Animado [The Crow and the Raven] Escrito por Nicholas Rubatino	01:42	Negativa	Positiva	Miedo / Desazón
Video casero creado con el fin de asustar al espectador con un susto.	00:23	Neutro	Positivo	Sobresalto / Sorpresa
Video musical [Bleed --- timmy trumpet & scndl] Humorístico con tono jocoso. Puesto como contrapunto al video anterior.	00:30	Positiva	Positivo	Regocijo
Video Musical [Around the World --- Daft Punk] con video del personaje Skeletor (Parte de un anuncio comercial)	01:04	Positiva	Positivo	Excitación / Disfrute
Video Musical de [Carpenter Brut - Roller Mobster] con escena de acción.	00:45	Positiva	Positivo	Excitación
Video Musical de [N'to - Trauma (Worakls Remix)] escenificando viaje en motocicleta.	01:00	Positiva	Positivo	Excitación / alegría
Video musical animado de:[ResuRection (Pacific Wave Mix) --- Dj Kharma] escenificando al personaje Bongo Cat 3d	00:52	Positiva	Positivo	Alegría
Video casero donde aparece niño pequeño jugando con su madre mientras suena la música (alegre) de [SUBWAY SURFERS 2020 CHICAGO]	00:11	Positiva	Positivo	Alegría
Fragmento de video humorístico	00:32	Positiva	Positivo	Alegría

Fragmento de video humorístico	01:00	Positiva	Positivo	Alegría
Video musical animado de [Matthew Wilder - Break My Stride]	00:26	Positiva	Positivo / Neutro	Alegría / Calma
Fragmento de video humorístico	00:32	Positiva	Positivo	Alegría
Fragmento de video humorístico	00:11	Positiva	Positivo	Alegría
Video casero de padre despertando a su hijo y jugando.	00:40	Positiva	Positivo / Neutro	Alegría / Felicidad
Video casero de baile de graduación donde suena la música [Buttercup · Jack Stauber]	00:23	Positiva	Positivo / Neutro	Alegría / Felicidad
Anuncio animado de Turismo de Oregon. Se muestran animaciones de gran calidad con música relajante	01:30	Positiva	Negativa / Neutro	Calma
Video de cocina	01:37	Neutra	Neutra	Calma / Aburrimiento.
Video animado con música tranquila	00:26	Neutra	Neutra	Calma / Aburrimiento.

Tabla 3.2 - Listado de videos utilizados en pruebas

En el capítulo de resultados veremos que efectivamente las emociones provocadas en los usuarios cubren todo el espacio de representación de las emociones.

3.4.2. Cuestionarios

Al finalizar la visualización de videos, realizamos a cada usuario un cuestionario para saber qué emociones experimentó a lo largo del video. Para ello utilizamos el llamado Self Assessment Manikin (SAM) questionnaire. [7] [6] [23]

Consiste en un cuestionario compuesto por monigotes en 3 escalas distintas: Pleasure, Valence, y Dominance. Si bien estas dimensiones se corresponden con las del modelo PAD de emociones, son compatibles con las del modelo circunflejo también. En cada escala hay varios grados en el rango [-1,1], y los monigotes representan cada uno de los grados. [7] [6]

Para expresar su emoción, el usuario elije uno de los monigotes en cada dimensión correspondiente a un grado. El conjunto de las dimensiones representa una emoción.

En nuestro caso utilizamos unos cuestionarios de 9 grados con saltos de 0.25.

Capítulo 3 - Experimentación

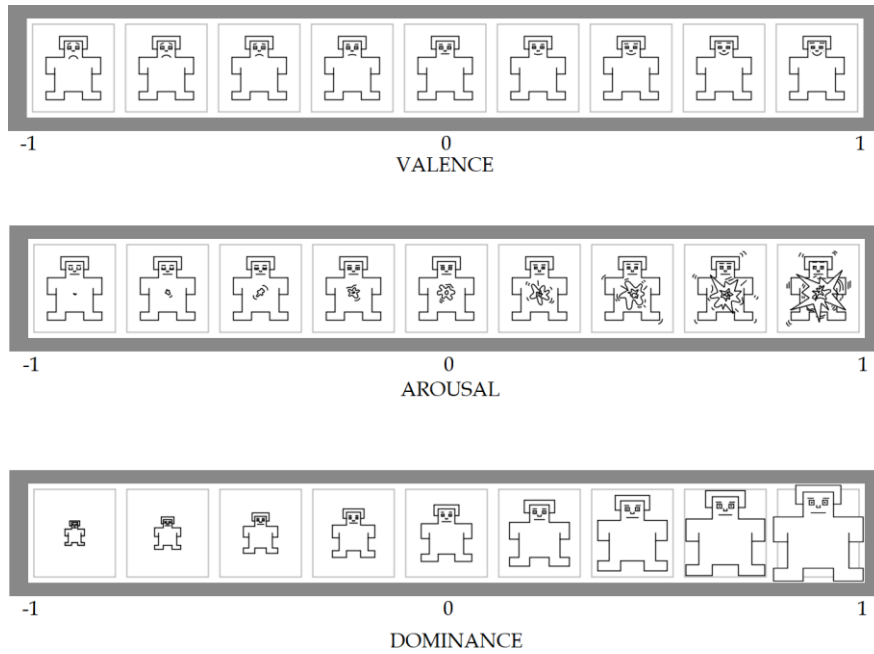


Ilustración 3.15 - Imágenes utilizadas en el Self Assessment Manikin questionnaire para medir valores de Arousal, Valence y Dominance

Existen formas alternativas de recopilar la respuesta emotiva por parte del usuario ante distintos eventos. Una de ellas es utilizar herramientas gráficas e interfaces como “EmojiGrid” (ver Ilustración 3.16) donde el usuario puede representar en cada momento la emoción que siente. Al mismo tiempo que el usuario está experimentando una emoción, en una escala gráfica en dos dimensiones que representan Arousal y Valence, puede señalar el punto justo donde se encuentra de forma sencilla. Esto permite además tener métricas continuas de la emoción sentida por el usuario en cada momento en contraposición a la emoción media principal que se siente en un evento extraída a través de cuestionarios. [66] [67]

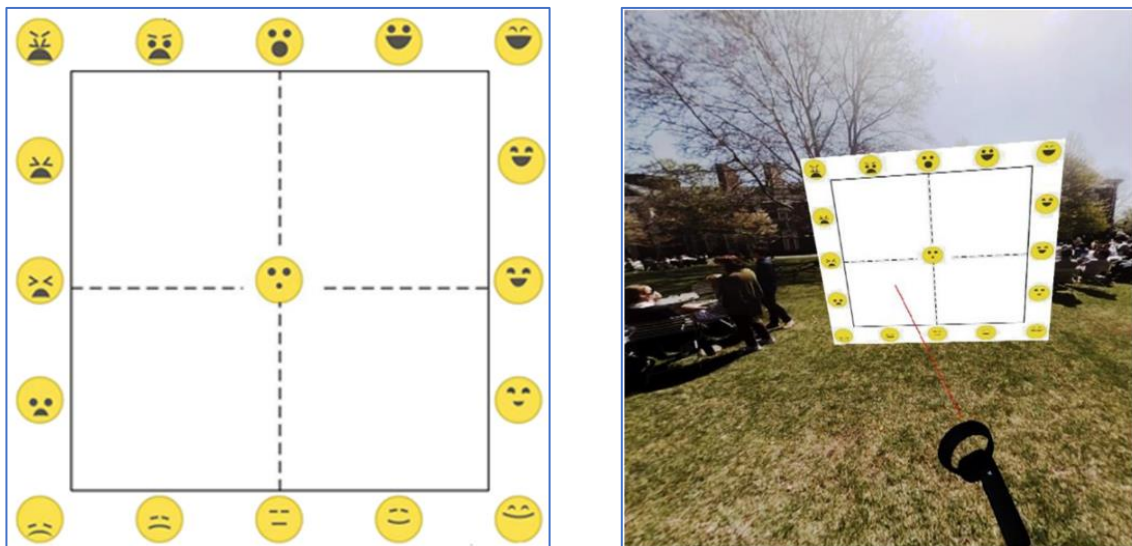


Ilustración 3.16 - Uso de EmojiGrid para evaluar emociones en escenarios de RV

No obstante, nosotros hemos preferido evitar este método por 2 motivos:

- Durante la realización de pruebas en escenarios virtuales no sería posible dado que una mano está ocupada con un controlador para moverse y la otra no está disponible por tener sensores puestos.
- Si el usuario se centra en dibujar, no está inmerso en el video y es más difícil provocarle una emoción real.

Aunque hemos añadido la dimensión “Dominance” a las métricas recopiladas, no es una dimensión a la que le daremos demasiada importancia.

3.4.3. Metodología de las pruebas

Las pruebas para crear el dataset que correlaciona señales fisiológicas con emociones mediante el visionado de videos se realizó en las siguientes condiciones:

- Cada prueba se desarrolló de forma individual en una habitación bien iluminada y ventilada donde solo se encuentra un voluntario (y el realizador de la prueba).
- El voluntario se coloca unos auriculares con supresión de ruido para que no le afecten sonidos externos.
- Antes de comenzar el visionado de videos, al voluntario se le colocan los sensores. En nuestro caso y debido a que no disponíamos de mejores sensores, 3 sensores en la mano derecha: 2 para medir el GSR, y otro para la frecuencia cardíaca.
- Una vez tenemos la captura de datos lista usando Arduino y un programa para capturar la entrada al *serial port* comenzamos la reproducción de los videos y dejamos solo al usuario para que no le afecte nuestra presencia.
- La duración total de todos los videos es de 35 minutos. Pasado ese tiempo volvemos con el usuario y le retiramos los sensores.
- Finalmente realizamos el cuestionario SAM al usuario.

Cada prueba tomó alrededor de 60 minutos en realizar sumando el tiempo de visualización, colocación de sensores, y realización del cuestionario con las pertinentes explicaciones al usuario. Realizamos pruebas a un total de 5 usuarios en días distintos. No obstante, tuvimos problemas con los sensores cuya fiabilidad nos gustaría recalcar, no era óptima. Tras comprobar los datos recopilados encontramos que los datos pertenecientes a 2 usuarios no eran válidos. Uno de ellos por datos corruptos (debido a bytes basura en la comunicación por *serial port* de Arduino con el ordenador), otro por un mal funcionamiento del sensor de frecuencia cardíaca que tomó valores medios de <10 pulsaciones por minuto. Esto supuso que finalmente contáramos únicamente con 3 conjuntos de datos válidos correspondiente a 3 usuarios distintos para formar nuestro dataset final.

3.4.4. Pruebas en entornos de RV

Debido a los sensores disponibles, durante las pruebas el usuario tiene una mano ocupada. Teniendo en cuenta esto, adaptamos los escenarios de RV para que pudieran controlarse con un único controlador. No obstante, al estar los sensores cableados mediante Arduino, el movimiento del usuario está limitado a mover la cabeza, ya que levantarse, girar el cuerpo, y mover demasiado los brazos suponen el riesgo de causar la mal función de los sensores.

Las pruebas consistieron someter al usuario a distintos eventos bien diferenciados:

Visualización de obras

En el entorno virtual es posible reproducir obras de teatro con actores que aparecen en el escenario (se aplica un efecto croma para que dé la sensación de que los actores están en el teatro y no es un video). Las obras están subtituladas.

Para este tipo de estímulo, se incluyeron 3 condiciones de test:

- Visualización de obra con audio máquina y subtulado plano: se reproduce una obra, pero en vez de reproducir la voz del actor, se reproduce sonido máquina sincronizado con los subtítulos creado artificialmente. Los subtítulos son blancos y no tienen ningún efecto especial.
- Visualización de obra con audio humano y subtulado plano: se reproduce una obra con la voz del actor. Los subtítulos son blancos y no tienen ningún efecto especial.
- Visualización de obra con audio humano y subtulado con estilo: se reproduce una obra con la voz del actor. Los subtítulos tienen varios estilos como distinto color en función de qué actor está hablando, efectos, etc.

Los videos reproducidos son distintos actos de la obra teatral “El caballero de Olmedo” que se grabaron con actores con la intención de ser usados en estos entornos virtuales. Durante la reproducción de los videos, el usuario es libre de mirar alrededor y moverse.

Puntos de interés

En el entorno virtual se añadieron elementos denominados puntos de interés. Estos consisten en puntos de información que resaltan elementos importantes del entorno virtual y a los que si el usuario se acerca proporcionan información sobre el mismo.

Para este tipo de estímulo, se consideraron 2 condiciones de test:

- Punto de información con información textual: la información es proporcionada en forma de texto flotante
- Punto de información con información auditiva: la información es dictada por audio máquina.

3.5. Exploración y Preprocesado de los datos

En este apartado explicaremos los datos con lo que trabajamos finalmente y como los preprocesamos para poder trabajar con ellos.

3.5.1. Exploración

Los datos con los que contamos finalmente tras realizar pruebas con usuario fueron:

- BPM (latidos por minuto) y GSR con un muestreo temporal irregular debido a limitaciones de la placa Arduino y los propios sensores.
- Un valor de Valence, Arousal, y Dominance por cada video.

De las 5 pruebas completas realizadas a usuarios tuvimos que descartar los datos tomados a 2 usuarios debido a corrupción de los datos.

De los 3 conjuntos de datos restantes encontramos problemas adicionales como pequeños rangos de tiempo del orden de varios segundos donde los sensores se congelaron y nos dejaron sin métricas. Este problema tuvimos que resolverlo interpolando para poder utilizar el resto de los datos viables.

Respecto a la distribución de las emociones, el dataset está aceptablemente balanceado. En la siguiente imagen vemos la cantidad de frames correspondientes a cada cuadrante del modelo de representación según respuestas de los usuarios:

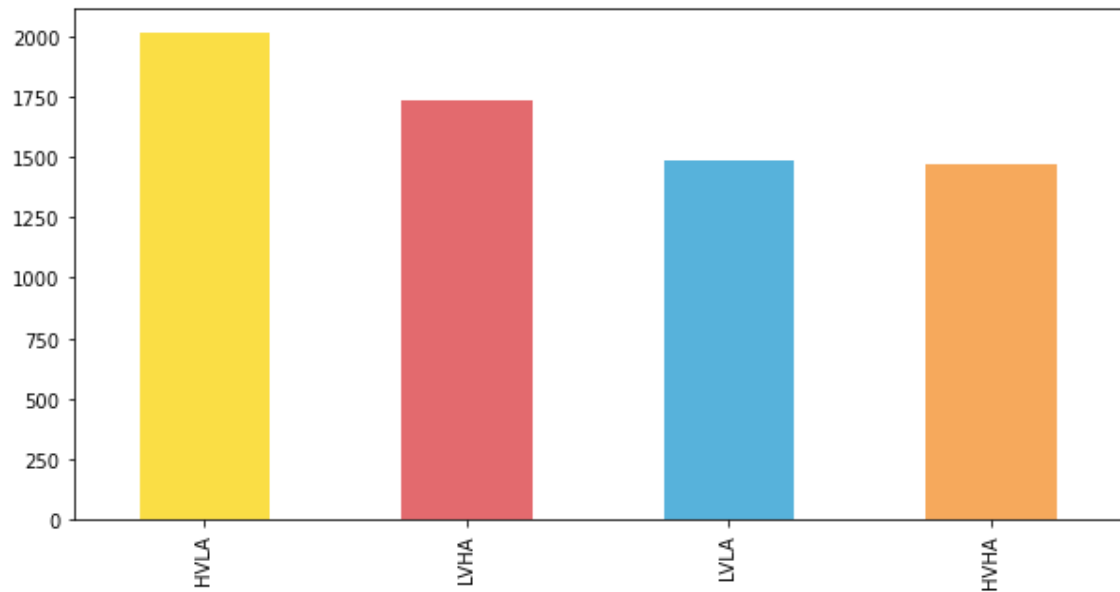


Ilustración 3.17 - Distribución de los videos por cuadrante emocional

3.5.2. Preprocesado

A continuación, describimos los pasos realizados para el preprocesado final de los datos recopilados. Algunos pasos han surgido fruto de la experimentación para alcanzar mejores resultados finales.

Respecto a datos originales recopilados del serial port de Arduino:

- Eliminación de líneas corruptas
- Corrección de líneas rotas fruto de errores en el envío desde el serial port de Arduino.
- Utilizando marcas de tiempo de los videos, añadir a cada entrada de datos fisiológicos el video visionado en ese momento y los valores de Valence, Arousal y Dominance experimentados por cada usuario durante el visionado del mismo.
- Conversión a formato .csv más fácil de manejar en Python.

Respecto a datos limpios en formato .csv:

- Lectura de archivos .csv y conversión a formato dataframe (Pandas) para facilitar las operaciones numéricas.
- Eliminación de outliers en datos fisiológicos. Por ejemplo, marcas de <10 pulsaciones por minuto
- Regularización de marcas temporales: los datos obtenidos de los sensores son irregulares en el tiempo, en algunos casos obtenemos 2 o 3 muestras por segundo y en otros una muestra cada par de segundos. Regularizamos a una muestra por segundo para tener datos coherentes.
- Rellenamos valores nulos fruto de la regularización temporal mediante interpolación de valores numéricos y uso del valor no nulo más cercano en caso de valores no numéricos.
- Cálculo de valores medios fisiológicos por cada usuario.

- Cálculo de datos adicionales como distancia al valor medio para cada usuario en BPM y GSR.
- Cálculo de diferencia en valores de datos fisiológicos entre muestras a 1, 5, y 10 segundos.
- A partir de valores de Valence y Arousal, añadimos el cuadrante en el modelo circunflejo de Russell correspondiente a cada muestra.
- Interpolación de valores de Valence, Arousal y Dominance en los cortes entre video al inicio de cada uno. Debido a que tenemos un único valor de Valence, Arousal y Dominance para cada video, para evitar saltos bruscos en valores la interpolación es necesaria y mejora los resultados finales. En la Ilustración 3.18 podemos ver a qué nos referimos con los saltos al observar valores de valence.

No hemos realizado “data augmentation” porque contamos únicamente con 2 métricas para estimar las emociones: HeartRate y GSR. Si intentamos aplicar data augmentation es posible que metamos más ruido al dataset y la precisión baje. En nuestro caso utilizaremos únicamente los datos recopilados.

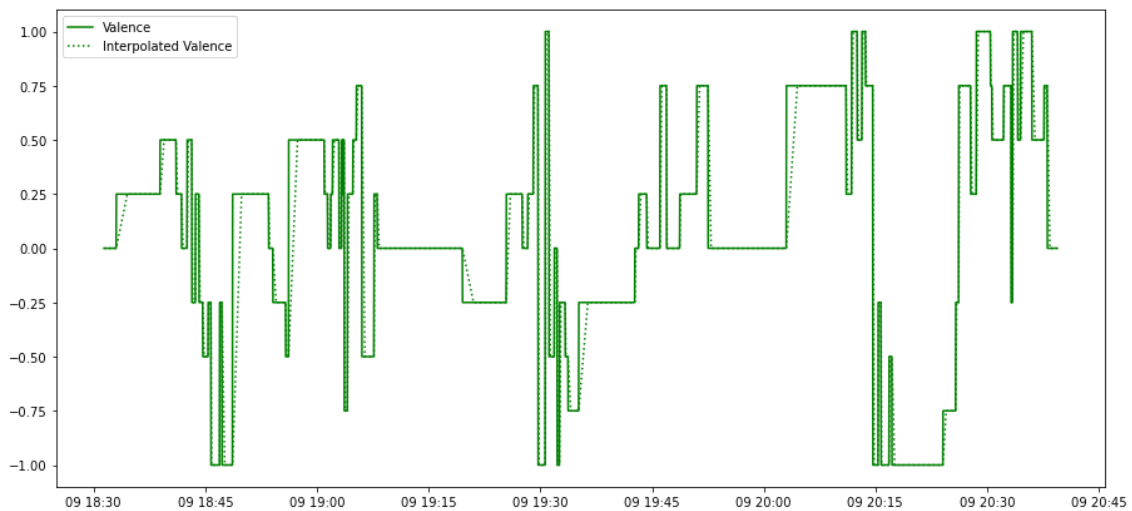


Ilustración 3.18 - Valores de Valence originales e interpolados

3.6. Creación y entrenamiento del modelo LSTM

Una red Long Short-Term Memory (LSTM) es un tipo de RNN especial que trabaja muy bien con problemas de predicción de secuencias. En nuestro caso trabajamos con datos secuenciales, en orden cronológico y que dependen en gran medida de estados anteriores, por lo que el modelo LSTM parece el idóneo para nuestro problema.

Para nuestro dataset hemos decidido utilizar una arquitectura de redes neuronales consistente en un modelo LSTM dado que es un tipo de RNN especial que trabaja muy bien con problemas de predicción de secuencias. Como ya hemos comentado anteriormente, al trabajar con datos secuenciales, en orden cronológico y que dependen en gran medida de estados anteriores, el modelo LSTM parece idóneo para nuestro problema. A continuación, cómo lo hemos creado, y cómo lo hemos entrenado.

3.6.1. Métricas utilizadas en experimentos

Nuestro objetivo es predecir qué emoción está experimentando el usuario en cada momento a partir de los datos fisiológicos que recopilamos. Para saber con qué grado de precisión nuestro modelo está funcionando necesitamos elegir una métrica adecuada.

Clasificación por cuadrante

En un primer lugar tratamos de ver si nuestro modelo aprendía a predecir el cuadrante de la emoción sentida por el usuario en el modelo circunflejo de Russell.

En este tipo de clasificación utilizamos las métricas de Accuracy, Recall y F1. Además, utilizamos matrices de confusión para observar los resultados de forma gráfica.

Un problema que encontramos es que hay muchos videos que caen en los límites entre cuadrantes y de forma arbitraria se asigna a un cuadrante u otro, repercutiendo en los resultados. Así nuestra predicción puede acercarse al punto real del espacio de representación, pero si el punto real está en un límite entre cuadrantes y nuestra predicción cae en el cuadrante equivocado, contará como un fallo.

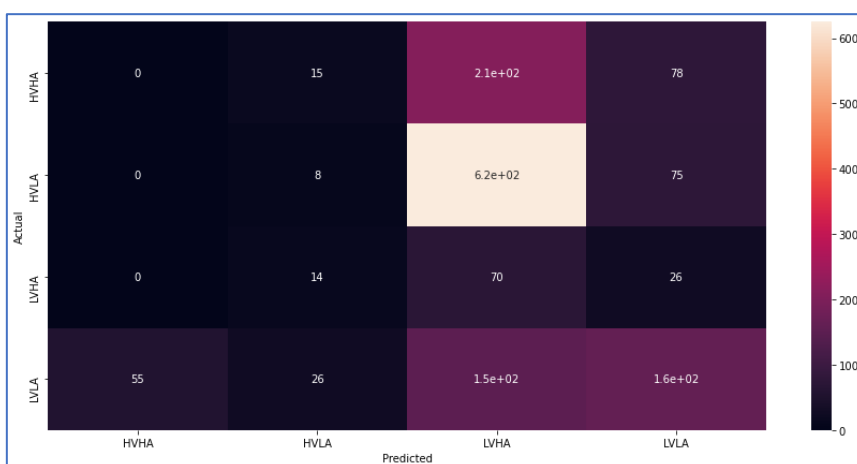


Ilustración 3.19 - Matriz de confusión

Distancia en el modelo de representación de las emociones

Puesto que lo que estamos intentando es localizar una emoción en el espacio de representación de Russell, decidimos probar como métrica la distancia euclídea entre nuestra predicción y la emoción real declarada por el usuario. Como trabajamos con secuencias temporales utilizamos la media de todas las predicciones y la última predicción.

Para saber la bondad de la predicción también tenemos en cuenta las características del espacio de representación utilizado. El rango es de $[-1,1]$ en ambas dimensiones, por lo que el máximo error teórico posible es de $2\sqrt{2}$, es decir, 2.82 unidades. Este caso se daría cuando la emoción real se encuentra en una esquina del espacio de representación y nuestra predicción se encuentra en la esquina contraria.

Para mostrar de forma más clara esta forma de evaluación, en la Ilustración 3.20 podemos ver un ejemplo extraído de los resultados finales. Podemos apreciar que a pesar de que las predicciones caen en cuadrantes distintos a los puntos reales, las predicciones y las emociones reales están relativamente cerca con una distancia media de 0.34 unidades.

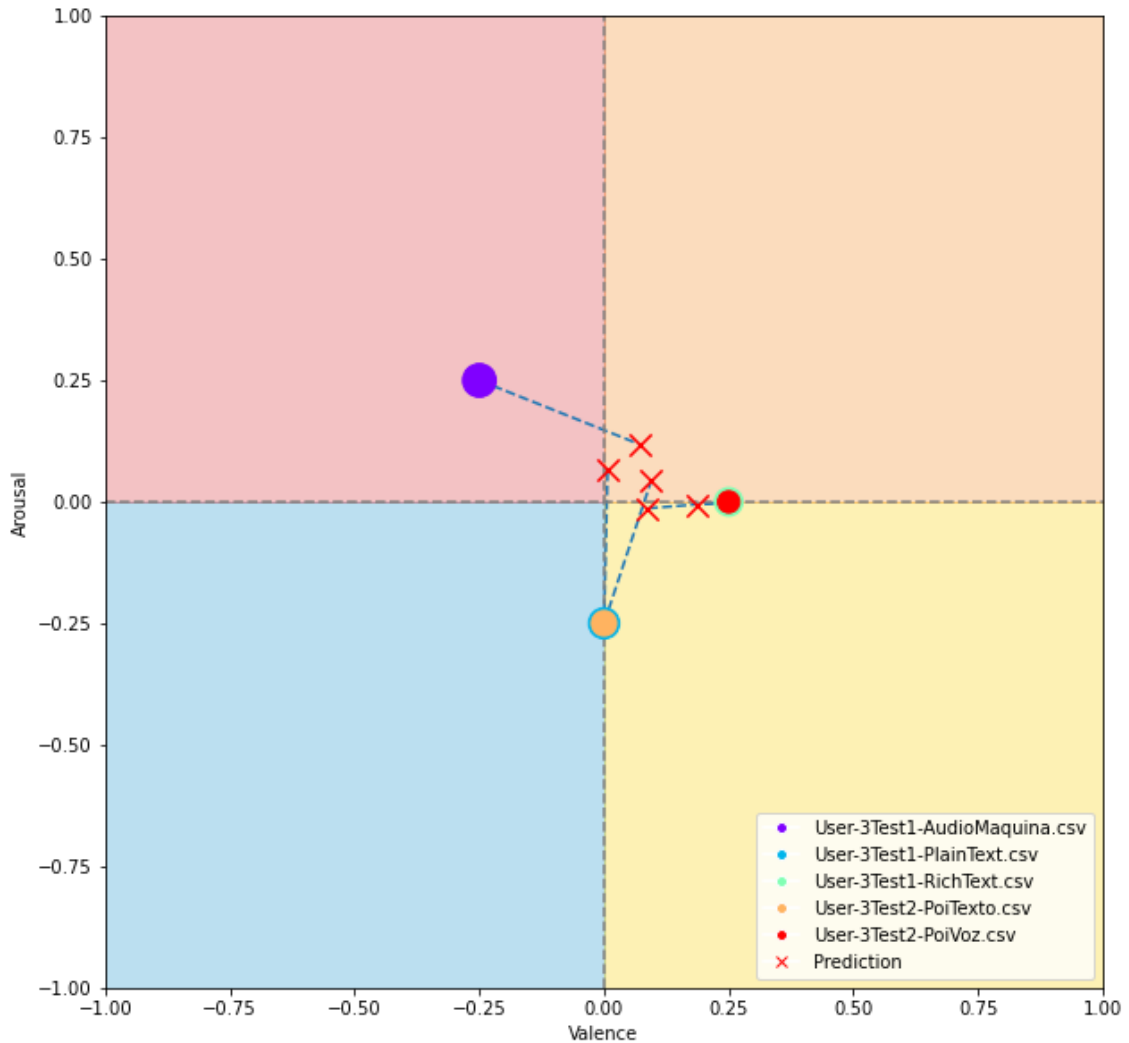


Ilustración 3.20 - Distancia entre predicciones y emociones reales. Tests del usuario 3.

3.6.2. División del dataset para realizar experimentos

La división del dataset en conjunto de entrenamiento y validación para realizar experimentos resultó un reto por 2 razones:

- Contamos con un número muy reducido de datos, con métricas de solo 3 usuarios distintos.
- Trabajamos con secuencias temporales, no podemos dividir los datos de cualquier forma.

Para realizar los experimentos probamos varias formas de dividir los datos. Las formas tradicionales como 10-fold cross-validation o leave-one-out cross-validation no nos resultaron adecuadas por no tener en cuenta los cortes entre videos.

Opción 1. División 70/30 aproximada

A partir de las marcas de tiempo de los videos y teniendo en cuenta que el orden de visionado fue el mismo en todos los usuarios, seleccionamos para el conjunto de entrenamiento las muestras correspondientes al conjunto de videos iniciales que suponían cerca del 70% de la lista de reproducción, y para evaluación el resto de los videos. No partimos ningún video por la mitad.

Esta forma de dividir los datos dio los peores resultados de todas las probadas. Creemos que es debido a que los videos están agrupados por emociones y los últimos videos corresponden a un mismo tipo de emoción que no ha sido entrenado en el primer 70% de las muestras.

Opción 2. División 2/1

En este caso probamos a utilizar dos usuarios para el conjunto de entrenamiento y un tercero para el conjunto de validación. El problema que encontramos fue que cada usuario era muy distinto y había mucha variabilidad dependiendo de qué usuario se eligiese para validación.

Opción 3. División mezclada de videos y usuarios

En este caso creamos el conjunto de validación eligiendo un video de cada alternando usuarios, y el resto de las muestras las utilizamos como conjunto de entrenamiento.

Esta forma de dividir los datos nos dio buenos resultados pues la variabilidad es baja independientemente de qué usuario utilices como comienzo. Asimismo, se utilizan todos los usuarios tanto para entrenamiento como para evaluación.

Por tanto, fue la opción seleccionada.

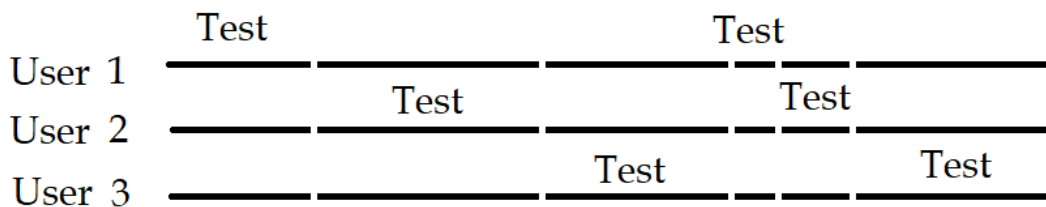


Ilustración 3.21 - Partición de dataset mezclando videos y usuarios

3.6.3. Experimentos

A continuación, listamos algunos de los experimentos que realizamos para decidir los mejores parámetros posibles a utilizar en el modelo de regresión final.

Para todos los experimentos utilizamos 50 épocas y utilizamos una parte de los datos como entrenamiento y otra como validación. En el modelo final utilizamos todos los datos para entrenamiento.

Mostramos los resultados en una tabla resumen al final de este apartado.

Como modelo de clasificación de cuadrantes

En un primer momento para probar distintas formas de dividir los datos en conjunto de entrenamiento y validación, utilizamos el modelo como clasificador de cuadrantes y no como regresor.

Los mejores resultados medios de precisión, recall y F1 se obtuvieron utilizando la división mezclada de videos y usuarios. En este experimento como es de esperar no pudimos aplicar la métrica de la distancia euclídea porque estamos usando el modelo como clasificador de cuadrante y no como regresor de valores de Valence y Arousal.

Interpolación

Probamos el modelo regresor interpolando los valores de Valence, Arousal y Dominance al inicio de cada video y sin interpolar. Esto recordamos que lo probamos para evitar “saltos” entre videos.

El resultado fue positivo, con interpolación obteníamos mejores resultados que se pueden apreciar en la tabla al final de este apartado.

Usar Dominance

Probamos haciendo regresión con Dominance o solo con Valence y Arousal. Como es de esperar, obtuvimos mejores resultados al descartar la dimensión de Dominance.

Nuestra hipótesis es que se debe a que Dominance es la dimensión peor definida por los usuarios al realizar los cuestionarios y a que es una dimensión secundaria con menor peso.

Escalado de valores numéricos

Hicimos pruebas escalando los distintos valores numéricos al rango $[-1,1]$ y sin escalar. Escalando obtuvimos resultados ligeramente mejores que se pueden observar en la tabla al final de este apartado.

Cálculo de diferencias numéricas entre frames

Probamos con los datos originales sin añadir columnas extra y añadiendo diferencias numéricas entre frames a 1, 5, y 10 segundos. En teoría, el modelo LSTM lo hace por sí solo. Sin embargo, obtuvimos resultados mejores al añadir diferencias temporales a los datos que se pueden consultar en la tabla al final de este apartado.

Funciones de pérdida

Comparamos las funciones de pérdida MAE (Mean Absolute Error) y MSE (Mean Square Error). Aunque no apreciamos diferencias significativas, finalmente decidimos utilizar la función MAE.

Función de activación

Puesto que estamos trabajando con un problema de regresión, nuestras opciones eran utilizar activación lineal o tangente hiperbólica. Los experimentos anteriores los hicimos todos con activación lineal. Al probar con la función de activación de tangente hiperbólica obtuvimos mejores resultados en nuestro caso.

Modelo

En todos los experimentos anteriores por pragmatismo y para reducir tiempo de entrenamiento utilizamos una estructura consistente en una capa LSTM de 50 neuronas y una capa de salida.

En este experimento probamos añadiendo capas LSTM y capas densas a la salida de éstas.

Finalmente nos quedamos con una arquitectura consistente en una capa LSTM de 128 unidades, otra capa LSTM de 64 unidades, una última capa LSTM de 32 unidades, una capa DENSE de 32 unidades con activación Relu, y una capa DENSE de salida con 2 neuronas y activación por tangente hiperbólica (tanh).

Resumen resultados experimentos

Capítulo 3 - Experimentación

En la siguiente tabla observamos los experimentos realizados en orden cronológico y separado por tipo de experimento. En cada experimento señalamos en negrita el mejor resultado obtenido y el parámetro asociado.

Div. Dataset	Dom.	Interpol.	Escalado	Diff temp	Loss	Activ,	Modelo	F1	Dist. Euc.
70/30	---	No	No	No	Categ. Cross.	softmax	LSTM50 Dense(4)	0.16	---
User 2/1	---	No	No	No	Categ. Cross.	softmax	LSTM50 Dense(4)	0.11	---
Mezcla	---	No	No	No	Categ. Cross.	softmax	LSTM50 Dense(4)	0.27	---
Mezcla	Si	No	No	No	MAE	Lineal	LSTM50 Dense(3)	0.28	1.01
Mezcla	Si	Si	No	No	MAE	Lineal	LSTM50 Dense(3)	0.29	0.97
Mezcla	no	no	No	No	MAE	Lineal	LSTM50 Dense(3)	0.29	0.73
Mezcla	no	si	No	No	MAE	Lineal	LSTM50 Dense(3)	0.29	0.71
Mezcla	no	si	si	no	MAE	Lineal	LSTM50 Dense(3)	0.34	0.67
Mezcla	no	si	si	Si (1)	MAE	Lineal	LSTM50 Dense(3)	0.32	0.67
Mezcla	no	si	no	Si (1)	MAE	Lineal	LSTM50 Dense(3)	0.32	0.67
Mezcla	no	si	si	Si 1,5,10	MAE	Lineal	LSTM50 Dense(3)	0.35	0.66
Mezcla	no	si	si	Si 1,5,10	MSE	Lineal	LSTM50 Dense(3)	0.33	0.69
Mezcla	no	si	si	Si 1,5,10	MAE	Lineal	LSTM64, LSTM32, DENSE 2	0.36	0.66
Mezcla	no	si	si	Si 1,5,10	MAE	Lineal	LSTM128, LSTM64, LSTM32, DENSE 32 DENSE 2	0.41	0.65
Mezcla	no	si	si	Si 1,5,10	MAE	Tanh	LSTM128, LSTM64, LSTM32, DENSE 32 DENSE 2	0.42	0.64

Tabla 3.3 - Resultados sobre el subconjunto de validación al realizar experimentos.

Las columnas de la tabla tienen el siguiente significado:

Div. Dataset: División del dataset

Dom. : Incluye dominancia o no.

Interpol: Interpolación de Valence, Arousal, Dominance al comienzo de los videos

Escalado: Escalado de valores numéricos al rango [-1,1]

Diff temp: inclusión de diferencias entre frames en datos.

Loss: función de pérdida usada

Activ.: Función de activación

Dist.Euc. : Distancia euclídea en modelo de representación.

Categ. Cross. : Categorical Cross-entropy (función de pérdida)

Tanh: Función de activación tangente hiperbólica.

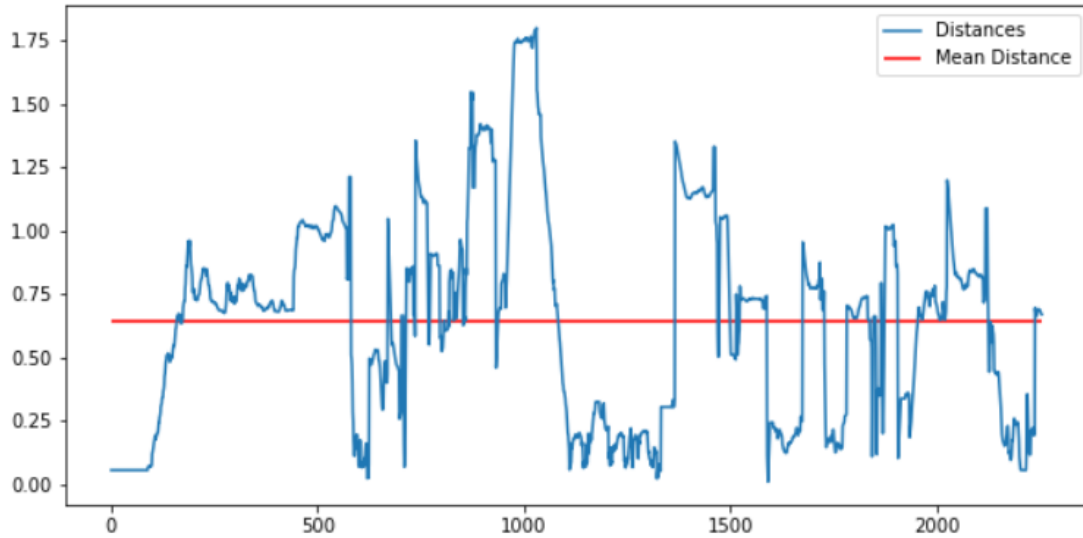
En la Ilustración 3.22 vemos los resultados de clasificación (a partir de los valores de Arousal y Valence calculados por regresión) del último modelo probado entrenado durante 50 épocas, sin dominancia, con interpolación, con escalado, con diferencias temporales, función de pérdida MAE, función de activación tanh, 3 capas LSTM (128, 64, 32), una capa densa de 32 (activación ReLU), y una capa densa de 2 neuronas. Observamos que obtenemos resultados de validación bajos con una accuracy global de 0.42:

	precision	recall	f1-score	support
HVHA	0.45	0.73	0.56	497
HVLA	0.40	0.96	0.56	596
LVHA	0.00	0.00	0.00	468
LVLA	0.00	0.00	0.00	694
accuracy			0.42	2255
macro avg	0.21	0.42	0.28	2255
weighted avg	0.20	0.42	0.27	2255

Ilustración 3.22 - Resultados clasificación último experimento

No obstante, a pesar de los aparentes malos resultados de clasificación por cuadrantes, en la Ilustración 3.23 podemos observar que la distancia euclídea media de nuestras predicciones a los puntos reales no es demasiado elevada (0.63 unidades).

Recordemos también que estos son resultados de experimentos, con modelos entrenados un número reducido de épocas y con un subconjunto de los datos, y testados con un subconjunto de validación. Los resultados no son significativos, solo nos sirven para determinar qué parámetros funcionan mejor.



Average Distance to Real point in model:
0.6460902515719442

Ilustración 3.23 - Resultados en distancia euclídea último experimento

3.6.4. Modelo Final

Nuestro modelo final consiste en 3 capas LSTM de 128, 64, y 32 unidades, una capa densa de 32 neuronas con función de activación ReLU, y una última capa densa de 2 neuronas con función de activación tanh. Las capas ocultas tienen dropout aplicado con un factor de 0.2. El modelo ha

lado entrenado durante 150 épocas. En la Ilustración 3.24 podemos ver un esquema de la estructura del modelo final entrenado.

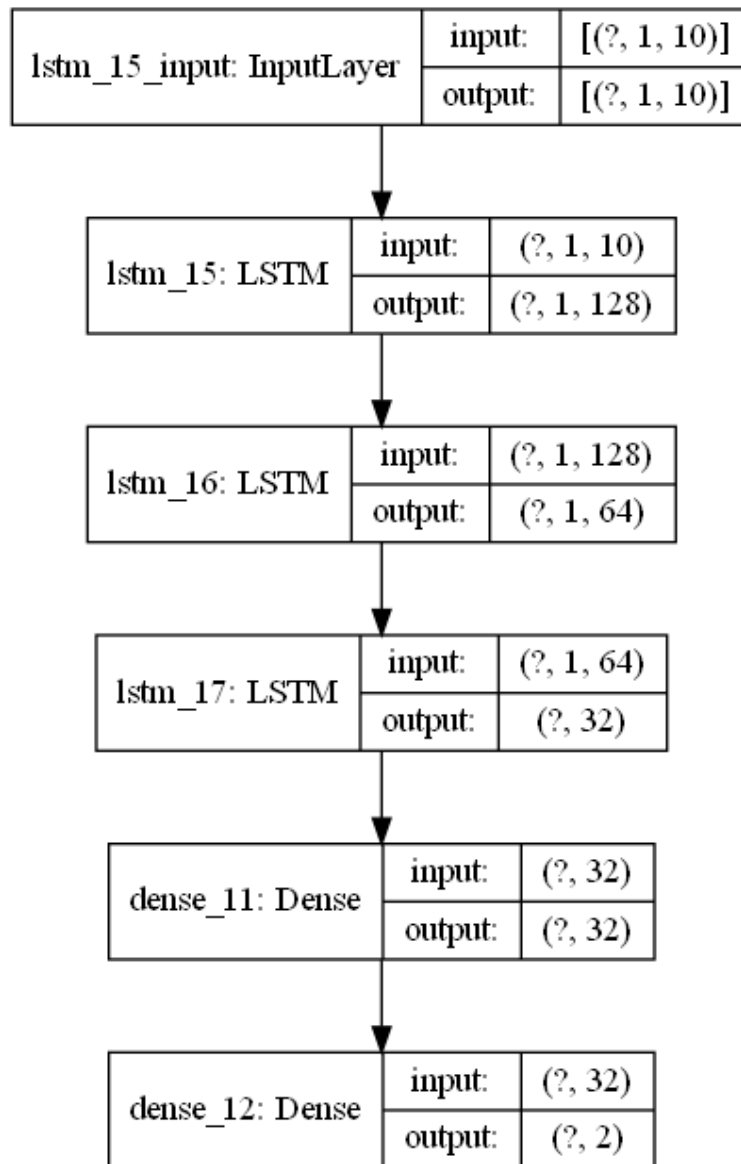


Ilustración 3.24 - Estructura modelo final entrenado

4. Resultados

A continuación, comentamos los resultados tanto para la creación del dataset propio, la creación del modelo de predicción de las emociones, como para la respuesta emotiva a los distintos eventos en entornos de RV.

4.1. Creación del dataset

Uno de nuestros objetivos era crear un dataset que relacionara emociones experimentadas por los usuarios con las métricas fisiológicas registradas a partir del cual crear nuestro modelo de regresión. Para crear dicho dataset seleccionamos una serie de videos con la intención de que provocasen una respuesta emocional en los usuarios diversa y que cubriese todo el espacio de representación de las emociones de Russell.

En la siguiente imagen mostramos la distribución de emociones sentidas por los usuarios que participaron en el estudio al visualizar los videos según sus respuestas. Podemos observar que, si bien cada usuario tiene su distribución característica, en general cubrimos todo el espacio de representación de las emociones.

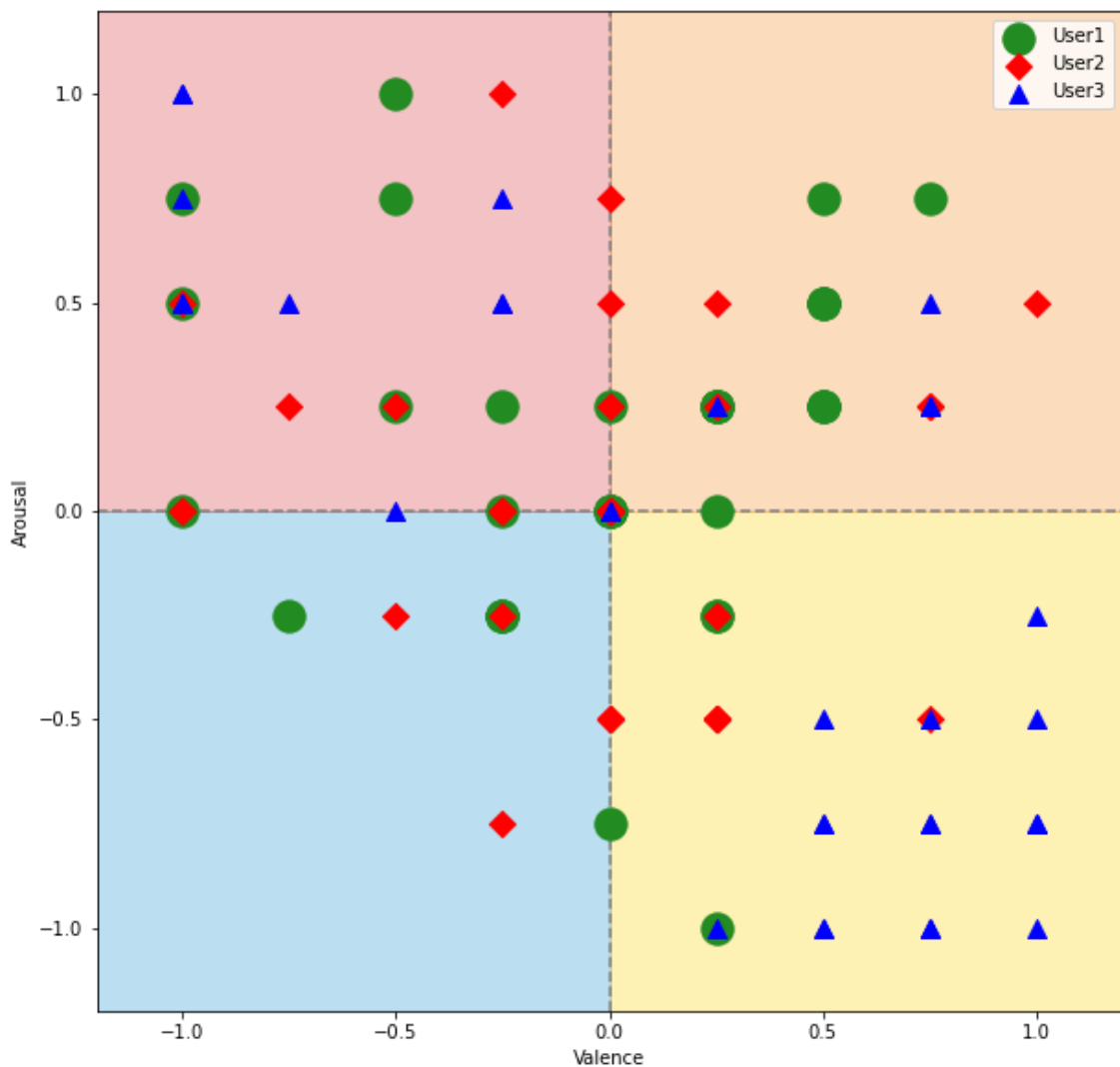


Ilustración 4.1 - Emociones provocadas por los videos en los usuarios

4.2. Resultados aplicados a Escenarios de RV

A partir de los datos reales obtenidos mediante cuestionarios a los usuarios sacamos los siguientes resultados respecto a la respuesta emotiva evocada por cada test.

Observaciones generales

En la imagen siguiente vemos las emociones sentidas por los usuarios en cada test, mediante sus respuestas reales. Podemos observar cómo independientemente de en qué cuadrante se encuentren, ningún evento ha producido una gran respuesta emotiva, todos los puntos se encuentran cerca del centro.

Esto nos indica que nuestros eventos no han sido muy eficaces para provocar emociones en el usuario.

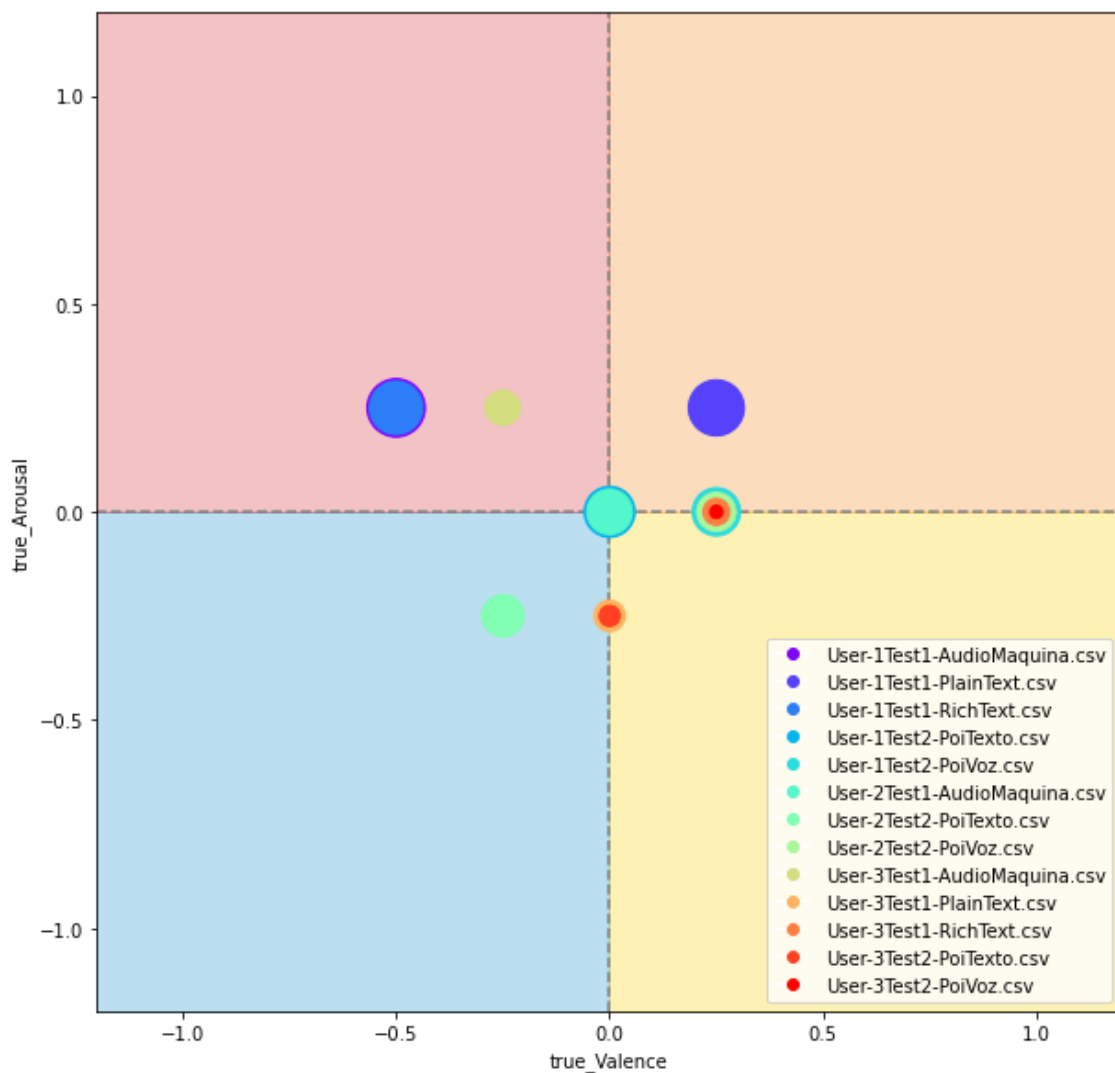


Ilustración 4.2 - Emociones provocadas por los tests en los usuarios

Visualización de obras

Visualización de obra con audio máquina y subtulado plano (Ilustración 4.3): es la que produce la respuesta emotiva más negativa, intensa y generalizada, con Valence negativa y Arousal positivo. Según las respuestas de los voluntarios, el hecho de ver al actor hablar mientras se escucha una voz producida artificialmente provoca un sentimiento de rechazo semejante al fenómeno del “Uncanny Valley”. Además, como la voz máquina va sincronizada con los subtítulos, en muchas ocasiones la voz y los gestos parecen desincronizados y aumenta la sensación de confusión.

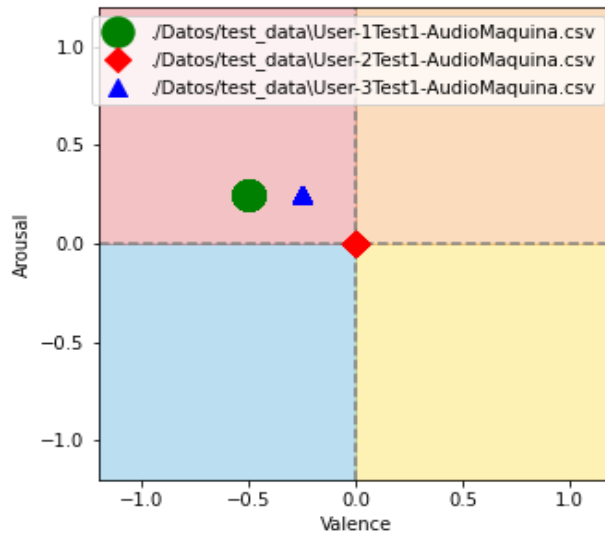


Ilustración 4.3 Resultados Reales - Test 1: Obra con audio máquina y texto plano

Visualización de obra con audio humano y subtulado plano (Ilustración 4.4): Produce una respuesta emotiva débil pero positiva. Obtenemos valores neutros o positivos de Valence. El Arousal depende del usuario. Según respuestas de los usuarios, el audio es adecuado (audio original grabado), y los subtítulos son fáciles de leer. No pudimos recopilar datos de uno de los usuarios por un fallo en los sensores.

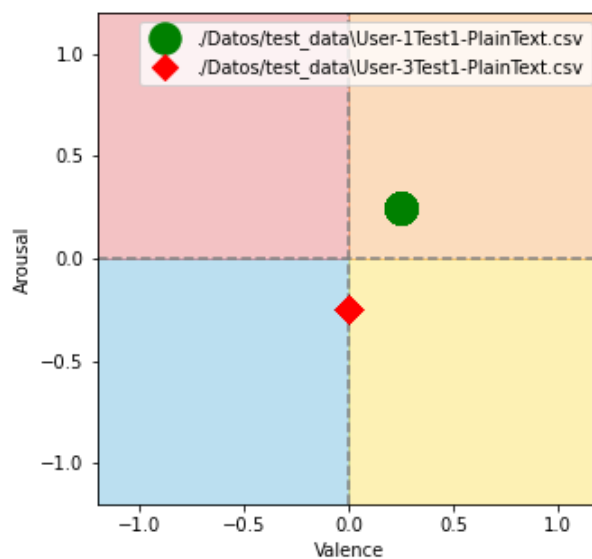


Ilustración 4.4 Resultados Reales - Test 1: Obra con audio humano y texto plano

- Visualización de obra con audio humano y subtulado con estilo (Ilustración 4.5): La respuesta emotiva depende del usuario en este caso. Mientras que para el usuario 3 es el evento multimodal que mejor respuesta emotiva provoca con Valencia positiva, para el usuario 1 es un evento que provoca el mismo rechazo que la obra con audio máquina. Según las respuestas de los usuarios esto se debe a simples preferencias personales: el usuario 1 prefiere texto simple mientras que el usuario 3 disfruta del uso de distintos estilos. No pudimos recopilar datos de uno de los usuarios por un fallo en los sensores.

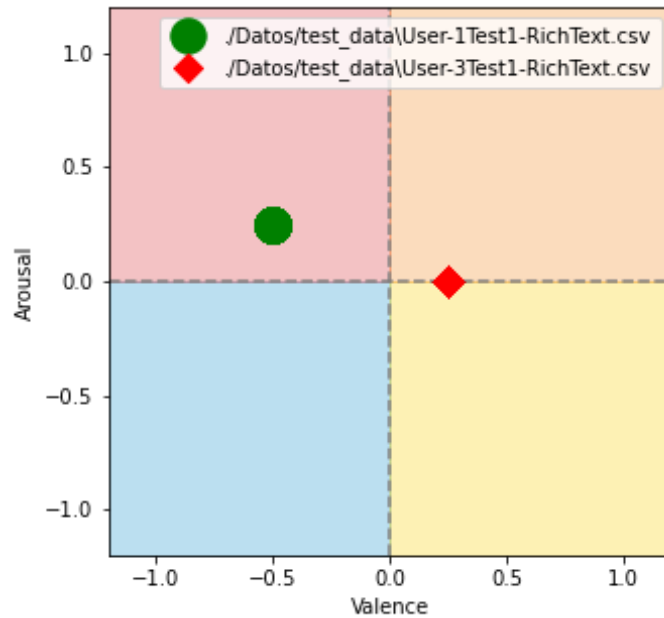


Ilustración 4.5 Resultados Reales - Test 1: Obra con audio humano y texto con estilo

Puntos de interés

El punto de interés basado en texto reporta una respuesta emotiva real más negativa que la del punto de interés basado en audio. El punto de interés basado en audio provoca en todos los usuarios una respuesta emotiva positiva y semejante.

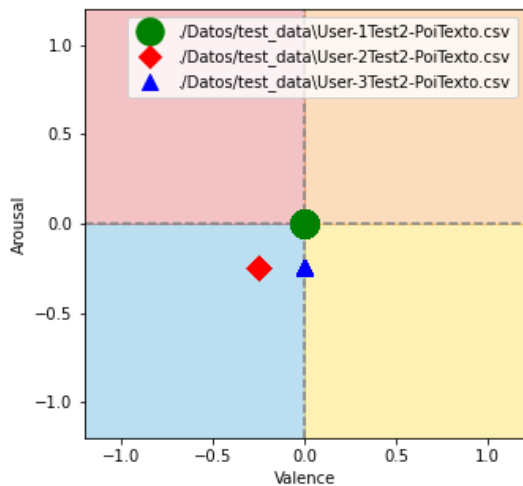


Ilustración 4.7 Resultados Reales - Test 2: Punto de Interés con Texto

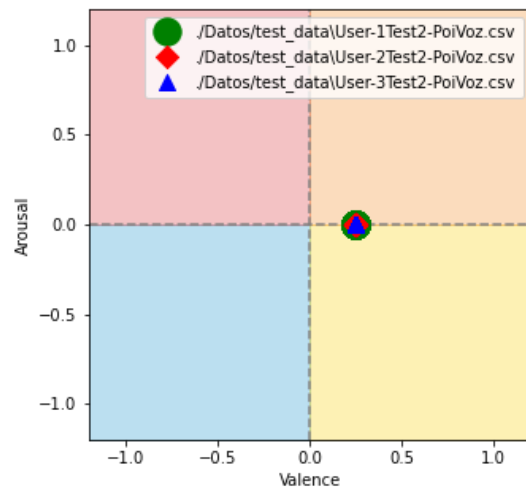


Ilustración 4.7 Resultados Reales - Test 2: Punto de Interés con Audio

4.3. Resultados del modelo de predicción de emociones

Basándonos en los resultados al realizar predicciones en los tests de usuario en entornos de RV podemos decir que hemos alcanzado unos resultados aceptables. Conseguimos una distancia euclídea media entre la predicción y el punto real en el espacio de representación de las emociones de 0.40, menos de la mitad del lado de un cuadrante.

Si trabajamos con emociones cercanas al centro del espacio de representación (emociones neutras) puede ser aún un mal resultado (debido a proximidad entre cuadrantes), pero con emociones intensas en los bordes del modelo se puede considerar buen resultado.

En nuestro caso, y dado que las emociones experimentadas por los usuarios al realizar los tests en los escenarios han caído normalmente cerca del centro y entre cuadrantes, nuestras predicciones caen en ocasiones en cuadrantes equivocados, pero a poca distancia de la emoción real.

A continuación, mostramos los resultados en el espacio de representación para todos los usuarios.

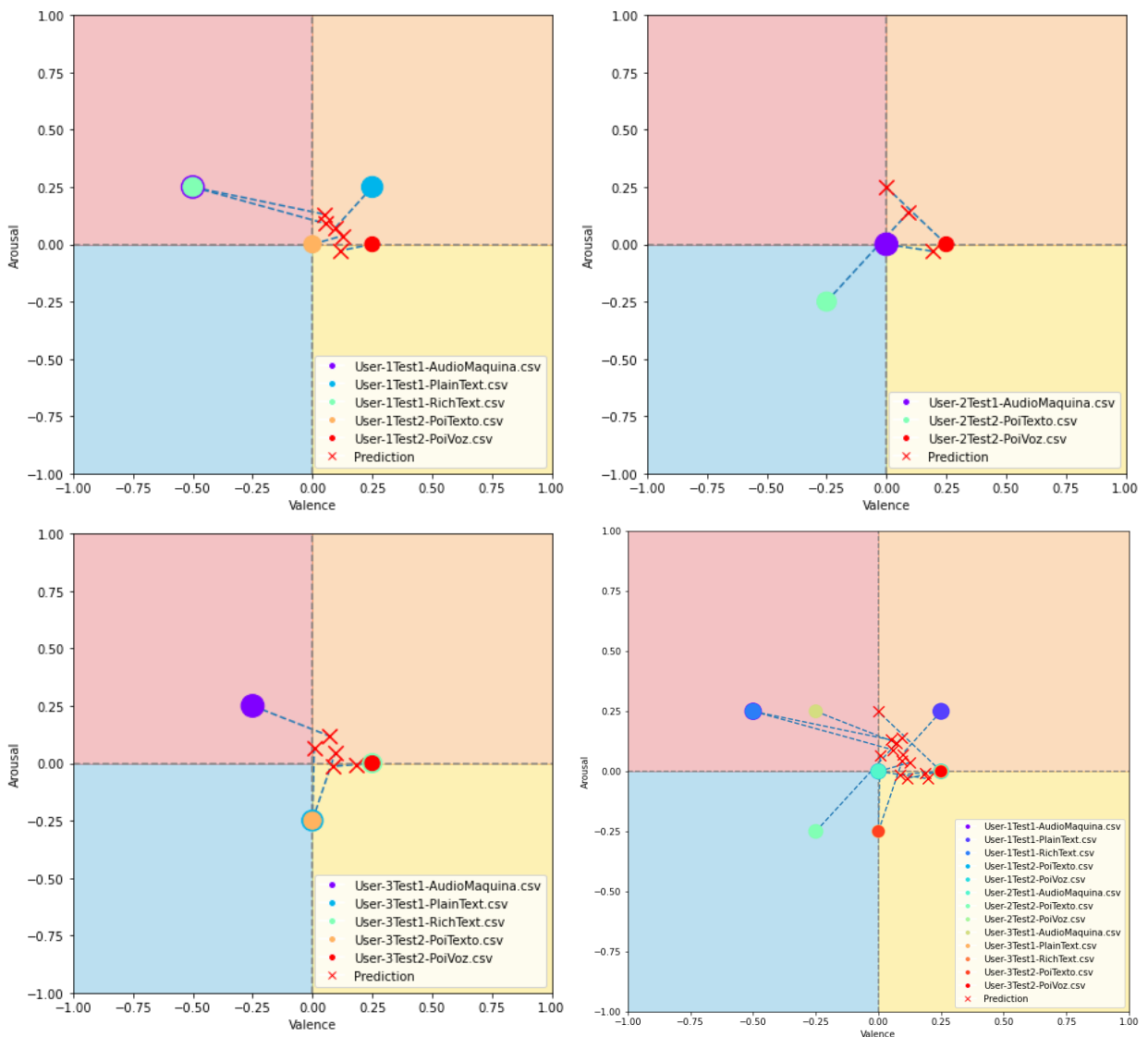


Ilustración 4.8 - Resultados de los tests en escenarios de RV por usuario y en conjunto

El resumen de los resultados es el siguiente:

Usuario	Test	Distancia Media Predicción Test	Distancia Media Predicción
Usuario 1	Vid. Audio Máquina Plain Text	0.635	0.440
	Vid. Audio Humano Plain Text	0.367	
	Vid. Audio Humano Rich Text	0.642	
	Point of Interest - Texto	0.314	
	Point of Interest - Audio	0.242	
Usuario 2	Vid. Audio Máquina Plain Text	0.282	0.446
	Point of Interest - Texto	0.618	
	Point of Interest - Audio	0.439	
Usuario 3	Vid. Audio Máquina Plain Text	0.454	0.342
	Vid. Audio Humano Plain Text	0.359	
	Vid. Audio Humano Rich Text	0.300	
	Point of Interest - Texto	0.338	
	Point of Interest - Audio	0.257	
Total Distancia Media		0.404	

Tabla 4.1 - Resumen resultados de los tests en escenarios de RV

5. Conclusiones y Trabajo Futuro

En este capítulo analizaremos los resultados alcanzados en el trabajo, extraeremos conclusiones y plantearemos posible trabajo futuro.

5.1. Análisis de los resultados

A continuación, analizamos los resultados que hemos alcanzado.

5.1.1. Calidad de los datos

A pesar del proceso de limpieza y preprocesado, los datos finales no ofrecen una calidad muy alta. Encontramos problemas a 3 niveles:

Escasez

Debido al tiempo requerido para cada prueba individual con usuario y a las pruebas que finalmente se tuvieron que descartar por corrupción de datos o problemas con sensores finalmente solo contamos con datos tomados a tres usuarios distintos.

Debido a los sensores disponibles solo se pudieron recopilar dos tipos de métricas: latidos por minuto (BPM) y valores de respuesta galvánica en la piel (GSR). Para una tarea compleja como la regresión en computación afectiva, métricas adicionales hubieran sido útiles.

Los datos finales para la creación del modelo de regresión son por tanto bastante escasos y no del todo adecuados para la creación de un modelo de regresión.

Calidad de las métricas recopiladas

Los datos fisiológicos recopilados tienen una fiabilidad limitada debido a la escasa calidad de los sensores utilizados. Los sensores utilizados no son médicos ni posiblemente de calidad suficiente para ser admitidos en estudios científicos, sino sensores asequibles adquiridos tras fallar otras alternativas.

Estos sensores además de proporcionar métricas de dudosa veracidad tuvieron el problema de que fallaron durante varios segundos durante la realización de las pruebas. A pesar de la interpolación aplicada en estos casos, esto afecta negativamente a los resultados. En la Ilustración 5.1 podemos ver valores de BPM a lo largo del tiempo. Las zonas planas son fruto de fallos en los sensores.

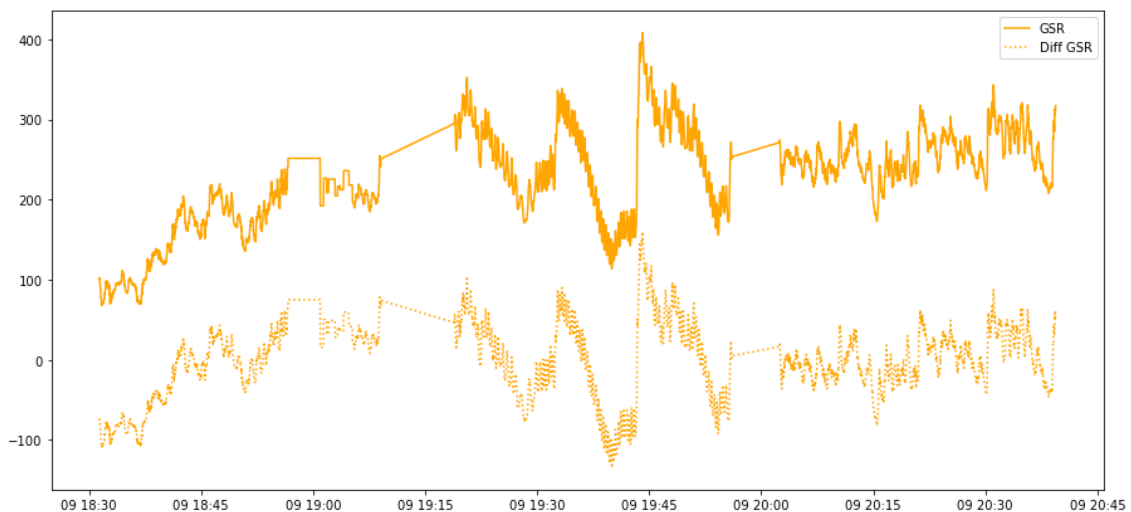


Ilustración 5.1 - Valores de GSR

Veracidad de las respuestas proporcionadas por los usuarios

Los valores de Valence, Arousal, y Dominance han sido recopiladas a partir de cuestionarios realizados a usuarios. A pesar de que a cada usuario se le ha explicado en qué consisten y cómo evaluarlas hasta estar razonablemente seguros de que han sido entendidas, no dejan de ser conceptos complejos.

En primer lugar, el usuario necesita auto analizar lo que siente con cada video, y puede malinterpretar sus sentimientos. Por ejemplo, puede sentir expectación ante un accidente inminente (por ejemplo, en el caso de un video donde aparece una mujer jugando con cuchillos) y malinterpretarlo como miedo.

Por otro lado, el usuario puede haber olvidado su sentimiento original al visionar un video. Los cuestionarios son rellenados después de visionar todos los videos y al evaluar cada uno individualmente el resto de los videos vistos puede influir en la percepción individual. Por ejemplo, si el usuario ve un video que le causa una sensación de aburrimiento en un primer momento, tras visionar otros videos que le causen desazón, miedo, rabia y otras emociones negativas, el usuario podría evaluar ese video como calmante en vez de aburrido. A pesar de advertir al usuario de estos errores, siguen siendo un riesgo.

Otro problema ya mencionado es la conversión de emociones definidas a valores de Valence, Arousal y Dominance. El usuario puede malinterpretar el cuestionario SAM y llegar a evaluar emociones como la calma con Arousal negativo con valores neutros (cero). La dimensión Dominance en especial ha dado bastantes problemas y finalmente la hemos descartado del modelo final puesto que además tampoco es tan significativa y no pertenece al modelo original de las emociones de Russell (el utilizado en este documento).

Finalmente, el usuario es capaz de mentir (no así los sensores por muy defectuosos que sean). Si bien no es un problema que hallamos encontrado con nuestros usuarios (presumiblemente), es un factor a tener en cuenta. Por ejemplo, si un usuario ve un video pensado para causar tristeza o compasión (por ejemplo, alguien teniendo un accidente) y realmente el usuario experimenta una emoción neutra o incluso de regocijo, el usuario puede verse tentado a mentir sobre la emoción experimentada para demostrar más empatía.

5.1.2. Tests en escenarios de RV

Uso de audio máquina VS uso de audio humano

El uso de audio máquina influye negativamente en la respuesta emotiva provocada en los usuarios causando una sensación de rechazo. El uso de audio humano (grabación original) produce una respuesta emotiva más positiva.

Uso de subtulado plano VS uso de subtulado con estilo

No queda claro qué tipo de subtulado provoca una respuesta emotiva mayor y más positiva dada la escasez de datos. En esta prueba contamos con 2 usuarios que dieron respuestas contradictorias. Concluimos con los datos disponibles que la respuesta emotiva provocada por cada uno de los modos de subtulado depende de preferencias personales de cada usuario, no obstante, no son conclusiones definitivas y nuevos datos podrían cambiarlas.

Uso de texto VS uso de audio en puntos de interés.

En este caso, el uso de audio para transmitir información de un punto de interés tiene una respuesta emotiva mayor y más positiva de forma generalizada. Una causa posible de este

fenómeno es el hecho de que utilizar audio en vez de texto otorga más libertad al usuario para moverse y mirar alrededor, cosa que no puede hacer mientras está leyendo un texto.

Otras cuestiones para considerar

Los voluntarios nos comentaron que sus emociones se podían haber visto influenciadas por el hecho de no poder moverse libremente por el entorno virtual dado que tenían una mano ocupada con sensores y que no podían moverse demasiado por estar conectados a los mismos.

Esto nos induce a pensar que, en circunstancias distintas, con sensores menos intrusivos y realizando los mismos tests, se hubieran obtenido resultados distintos con respuestas emotivas mayores.

Tampoco ayudan a crear una respuesta emotiva las obras mostradas pertenecientes a la novela “El caballero de Olmedo”. Los usuarios nos comentaron que no es una obra que les llamase la atención y les llegó a parecer aburrida.

5.2. Conclusiones

Respecto a resultados del modelo

- Los resultados del modelo de inferencia de emociones a partir de métricas fisiológicas han sido aceptables, aunque con mucha capacidad de mejora.
- Las métricas de precisión de clasificación en cuadrantes (F1, Accuracy, etc.) son bajas debido a muchas emociones cayendo en límites entre cuadrantes y la clasificación arbitraria en cuadrantes de dichas emociones entre cuadrantes.
- Estos resultados probablemente son debidos a las escasas métricas utilizadas para hacer la predicción, únicamente frecuencia cardíaca y GSR, las cuales pueden variar debido a distintas causas y por sí mismas pueden no ser tan significativas como pudieran ser otras más específicas.
- Otra causa de los malos resultados puede ser una baja precisión de los sensores utilizados. En muchas ocasiones las lecturas no eran correctas.

Respecto a eventos de escenarios de RV

- En entornos virtuales, el audio humano crea una mayor y más positiva respuesta emotiva, mientras que el audio máquina produce una respuesta negativa.
- Es más cómodo para el usuario el uso de puntos de interés con audio. La necesidad de leer texto limita la movilidad del usuario y puede tener una respuesta emotiva negativa.
- Limitaciones de movilidad (como el hecho de llevar sensores puestos) reducen la respuesta emotiva.

Conclusiones personales

Si bien se han alcanzado los objetivos propuestos, los resultados obtenidos son mejorables y se espera haber aprendido de los fallos cometidos para no repetirlos en un futuro. No obstante, teniendo en cuenta las limitaciones a las que nos hemos enfrentado (sensores de baja calidad, reducido número de usuarios, coste de cada prueba individual, fallos en sensores, etc.), estamos satisfechos con el trabajo realizado.

El campo de la computación afectiva unido al uso de entornos de RV es un campo con mucho potencial y está aún sin explotar. La realización de este trabajo ha servido al autor como introducción al campo de la computación afectiva y su integración con entornos de RV

multimodales. Se espera que el trabajo realizado sirva como base para abordar una futura tesis doctoral en este campo financiada por una beca FPU.

5.3. Problemas encontrados durante la realización del trabajo

- Dificultad para adquirir sensores que permitiesen adquirir y almacenar datos fisiológicos y los retrasos producidos por dicha razón.
- Dificultad para explicar la dimensión de la dominancia a los usuarios, si bien hay emociones donde está clara, hay otras donde es una dimensión confusa.
- Elevada duración de cada prueba individual: ~1,5h por usuario.
- La realización de pruebas se ha visto entorpecida por el COVID-19. Después de cada prueba con cada usuario es necesario desinfectar sensores, casco de realidad virtual, controladores, etc. Además, es complicado reunir a varias personas el mismo día por relucencia de los voluntarios a juntarse con desconocidos.
- El hecho de tener una mano ocupada en entornos de realidad virtual es un gran impedimento para experimentar la experiencia completa y eso se refleja en los resultados.
- Falta de datasets en el ámbito del reconocimiento de las emociones, que sean de libre acceso.
- Falta de un estándar para la realización de datasets del ámbito de reconocimiento de emociones. No solo hay muchas métricas posibles, sino que la forma de tomar dichas mediciones puede variar y dar lugar a resultados distintos.
- Se depende de las respuestas del usuario para crear un set de emociones “reales” experimentadas. Esto es un problema dado que:
 - o El usuario puede creer que está experimentando una emoción mientras siente otra y darnos resultados de forma errónea por equivocarse al interpretar su propia emoción.
 - o Puede dar resultados falsos por diversos motivos. Por ejemplo, puede ver dibujos animados, que le produzcan felicidad y por vergüenza reportar haber sentido calma o aburrimiento, que están en cuadrantes distintos del modelo circunflejo.
 - o El usuario puede no haber entendido las escalas y distintas dimensiones del modelo emocional utilizado y equivocarse al contestar.
- Factores externos pueden condicionar los datos fisiológicos de los usuarios, como por ejemplo el calor, la incomodidad, llevar un rato sentado, o llevar horas sin comer son factores que pueden variar la frecuencia cardíaca de alguien y varían los resultados.
- Cada persona siente de una forma distinta y videos para provocar una emoción pueden resultar inútiles o causar emociones no esperadas.

5.4. Trabajo futuro

Hay bastantes líneas de trabajo futuro en el ámbito de la computación afectiva. Por nombrar algunos ejemplos:

Datasets

Uno de los puntos que hemos recalado en este trabajo es la escasez de datasets de libre acceso para trabajar en el campo de la computación afectiva. Una de las posibles líneas de trabajo futuro sería analizar la necesidad de un estándar común para hacer para la toma de métricas en el campo de la computación afectiva.

Por supuesto también es necesario trabajar en la creación de datasets libres, con métricas estandarizadas con pruebas extensas y exhaustivas que sean útiles para diversos ámbitos.

Estudio de nuevos eventos multimodales

Como trabajo futuro, sería conveniente añadir nuevos estímulos a los eventos multimodales vistos en entornos de RV. Hablamos de estímulos multisensoriales como pueden ser el uso del olfato (a través de dispensadores de olor), o del tacto (a través de guantes hápticos). Sería interesante estudiar cómo estos nuevos estímulos incrementarían la sensación de inmersión virtual y analizar la variación en la respuesta emotiva provocada en el usuario.

Otra área para explorar sería la de la “holo presencia”, es decir, la recreación del cuerpo del usuario en el entorno virtual de forma hiperrealista, y estudiar la respuesta emotiva que provoca sobre todo durante la interacción con otros usuarios. [19]

Posibles aplicaciones futuras

La computación afectiva, los entornos de RV, y la combinación de ambos tienen muchas aplicaciones potenciales. Por poner algunos ejemplos podemos citar:

La creación de cástings automatizados: a partir de audio grabado por actores se pueden utilizar a usuarios y en base a sus reacciones y utilizando un modelo de regresión de emociones, determinar de forma objetiva qué actores logran una mayor respuesta emocional en los oyentes. Otra alternativa más ambiciosa sería a partir del audio, determinar de forma directa (sin usar usuarios) la emoción transmitida mediante distintos modelos. Para ello haría falta un sistema más complejo que por ejemplo fuera capaz de detectar el énfasis de la voz del actor, los matices, etc.

Aplicar de la computación afectiva para ayudar en el ámbito de la educación a adaptar la forma de aprendizaje y formación a las características individuales de cada persona, y servir como apoyo extra a personas con necesidades especiales como por ejemplo en el caso de niños autistas.

Bibliografía

- [1] [Online]. <https://blog.gft.com/es/2018/11/06/como-usar-redes-neuronales-lstm-en-la-prediccion-de-averias-en-las-maquinas/>
- [2] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, and Andy Davis. (2016) TensorFlow: A system for large-scale machine learning. [Online]. <https://www.tensorflow.org/>
- [3] Kylee Ack Baraly et al., "Database of Emotional Videos from Ottawa (DEVO)," *Collabra: Psychology*, vol. 6, p. 10, Feb. 2020.
- [4] Salma Alhagry, Aly Aly, and Reda El-Khoribi, "Emotion Recognition based on EEG using LSTM Recurrent Neural Network," *International Journal of Advanced Computer Science and Applications*, vol. 8, Oct. 2017.
- [5] Mohammed Almashraee, Dagmar Monett, and Adrian Paschke, "Emotion Level Sentiment Analysis: The Affective Opinion Evaluation," , May 2016.
- [6] Margaret M. Bradley and Peter J. Lang, "Measuring emotion: the Self-Assessment Manikin and the Semantic Differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, pp. 49-59, 1994.
- [7] Teah-Marie Bynion and Matthew Feldner, Self-Assessment Manikin, 2017.
- [8] Erion Çano and Maurizio Morisio, Music Mood Dataset Creation Based on Last FM Tags, May 2017, Department of Control and Computer Engineering, Polytechnic University of Turin.
- [9] Francisco Casacuberta, "Notas de clase de Machine Translation. Tema 3 - Neural Machine Translation," Universidad Politécnica de Valencia, Diciembre 16, 2020.
- [10] HongYi Chen, Kai Chin, and Chrystalle Tan, "Selection and validation of emotional videos: Dataset of professional and amateur videos that elicit basic emotions," *Data in Brief*, vol. 34, p. 106662, Feb. 2021.
- [11] François Chollet. (2015) keras. [Online]. <https://keras.io>
- [12] Lucio Ciabattone et al., Real-time mental stress detection based on smartwatch, Jan. 2017.
- [13] Weihui Dai, Dongmei Han, Yonghui Dai, and Dongrong Xu, "Emotion Recognition and Affective Computing on Vocal Social Media," *Information & Management*, vol. 52, Feb. 2015.
- [14] Najam Dar, Muhammad Akram, Sajid Khawaja, and Amit Pujari, "CNN and LSTM-Based Emotion Charting Using Physiological Signals," *Sensors*, vol. 20, p. 4551, Aug. 2020.
- [15] Francisco de Arriba Pérez, Manuel Caeiro Rodriguez, and Juan Santos, "Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios," *Sensors*, vol. 16, p. 1538, Sep. 2016.

- [16] Andrius Dzedzickis, Arturas Kaklauskas, and Vytautas Bučinskas, "Human Emotion Recognition: Review of Sensors and Methods," *Sensors*, vol. 20, p. 592, Jan. 2020.
- [17] Maria Egger, Matthias Ley, and Sten Hanke, "Emotion Recognition from Physiological Signal Analysis: A Review," *Electronic Notes in Theoretical Computer Science*, vol. 343, pp. 35-55, May 2019.
- [18] Michalis Feidakis et al., "Affective Assessment in Distance Learning: A Semi-explicit Approach," *International Journal of Technologies in Learning*, vol. 26, pp. 19-34, Apr. 2019.
- [19] S. Fernandez, Mario Montagud, G. Cernigliaro, and D. Rincón, "Toward Hyper-realistic and Interactive Social VR Experiences in Live TV Scenarios," *IEEE Transactions on Broadcasting*, To appear in 2021.
- [20] Nestor Garay-Vitoria, Idoia Cearreta, Juan López, and Inmaculada Fajardo, "Assistive Technology and Affective Mediation," *International Journal of Human Technology ISSN: 1795-6889*, vol. 2, pp. 55-83, Apr. 2006.
- [21] Vishnunarayan Girishan Prabhu, Courtney Linder, Laura Stanley, and Robert Morgan, "An Affective Computing in Virtual Reality Environments for Managing Surgical Pain and Anxiety," in *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2019, pp. 235-2351.
- [22] Martin Gjoreski et al., "Datasets for Cognitive Load Inference Using Wearable Sensors and Psychological Traits," *Applied Sciences*, vol. 10, p. 3843, May 2020.
- [23] Dini Handayan and Hamwira Yaacob, "Recognition of Emotions in Video Clips: The Self-Assessment Manikin Validation," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, p. 1343, Dec. 2015.
- [24] Neska Haouij, Biosignals for driver's stress level assessment : functional variable selection and fractal characterization, July 2018, Université Paris-Saclay; École nationale d'ingénieurs.
- [25] Neska Haouij, Jean-Michel Poggi, Sylvie Ghalila, Raja Ghozi, and Jaïdane Mériem, "AffectiveROAD System and Database to Assess Driver's Attention," *SAC '18: Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, Apr. 2018.
- [26] Charles R. Harris et al., "Array programming with NumPy.," *Nature*, vol. 585, pp. 357--362, 2020. [Online]. <https://numpy.org/>
- [27] André Henriksen et al., "Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables," *Journal of Medical Internet Research*, vol. 20, p. e110, Mar. 2018.
- [28] <https://www.olorama.com/es>.
- [29] John Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, vol. 9, pp. 90-95, June 2007.

- [30] Dong Pyo Jang et al., "Analysis of Physiological Response to Two Virtual Environments: Driving and Flying Simulation," *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, vol. 5, pp. 8-11, Mar. 2002.
- [31] Jayanta Kalita and Nabamita Deb, "Emotional Text to Speech Synthesis: A Review," *IJARCCCE*, vol. 6, pp. 428-430, Apr. 2017.
- [32] Arvid Kappas, Eva Krumhuber, and Dennis Küster, "Facial behavior," in *Handbook of Nonverbal Communication.*, 2013, pp. 131-166.
- [33] Stamos Katsigiannis and Naeem Ramzan, "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 98-107, Jan. 2018.
- [34] Mitu Khandaker-Kokoris, "Designing affective video games to support the social-emotional development of teenagers with Autism spectrum disorders," *Studies in health technology and informatics*, vol. 144, pp. 37-9, Feb. 2009.
- [35] Trinh Khanh, S.H. Kim, Guesang Lee, Hyung-Jeong Yang, and Eu-Tteum Baek, "Korean video dataset for emotion recognition in the wild," *Multimedia Tools and Applications*, vol. 80, pp. 1-14, Mar. 2021.
- [36] Sander Koelstra et al., "DEAP: A Database for Emotion Analysis Using Physiological Signals," *IEEE Transactions on Affective Computing*, vol. 3, pp. 18-31, Dec. 2011.
- [37] Agata Kołakowska, Agnieszka Landowska, Mariusz Szwoch, Wioleta Szwoch, and Michał Wróbel, *Modeling emotions for affect-aware applications.*, 2015.
- [38] (2021, Octubre) libelium. [Online]. <https://www.libelium.com/es/>
- [39] Leandro Mano et al., "An intelligent and generic approach for detecting human emotions: a case study with facial expressions," *Soft Computing*, vol. 24, June 2020.
- [40] Javier Marín-Morales et al., "Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors," *Scientific Reports*, vol. 8, Sep. 2018.
- [41] Javier Marín-Morales, Carmen Llinares, Jaime Guixeres, and Mariano Alcañiz Raya, "Emotion Recognition in Immersive Virtual Reality: From Statistics to Affective Computing," *Sensors*, vol. 20, pp. 51-63, Sep. 2020.
- [42] Wes Mckinney, "pandas: a Foundational Python Library for Data Analysis and Statistics," *Python High Performance Science Computer*, Jan. 2011.
- [43] Juan Miranda, Mojtaba Khomami Abadi, Nicu Sebe, and Ioannis Patras, "AMIGOS: A dataset for Mood, personality and affect research on Individuals and GrOupS," *IEEE Transactions on Affective Computing*, vol. PP, Feb. 2017.

- [44] Mohammadhossein Moghimi, Robert Stone, Pia Rotshtein, and Neil Cooke, "Influencing Human Affective Responses to Dynamic Virtual Environments," *Presence: Teleoperators and Virtual Environments*, vol. 25, pp. 81-107, Nov. 2016.
- [45] Mario Montagud, Cristian Hurtado, **Juan Antonio De Rus**, and Sergi Fernández, "Subtitling 3D VR Content with Limited 6DoF: Presentation Modes and Guiding Methods," *Applied Sciences*, vol. 11, p. 7472, Aug. 2021.
- [46] Mario Montagud, **Juan Rus**, Rafael Fayos-Jordan, Miguel Garcia-Pineda, and Jaume Segura Garcia, "Open-source software tools for measuring resources consumption and DASH metrics," in *MMSys '20: 11th ACM Multimedia Systems Conference*, 2020, pp. 261-266.
- [47] Mario Montagud, Jaume Segura Garcia, **Juan Antonio De Rus**, and Rafael Fayos-Jordan, "Towards an Immersive and Accessible Virtual Reconstruction of Theaters from the Early Modern: Bringing Back Cultural Heritage from the Past," in *IMX '20: ACM International Conference on Interactive Media Experiences*, June 2020, pp. 143-147.
- [48] Norma Ontiveros-Hernández, Miguel Pérez-Ramírez, and Yasmin Hernandez, "Virtual Reality and Affective Computing for Improving Learning," *Research in Computing Science*, vol. 65, pp. 121-131, Dec. 2013.
- [49] Fabian Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, Jan. 2012.
- [50] David Pollreis and Nima Taherinejad, "A simple algorithm for emotion recognition, using physiological signals of a smart watch," in *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2017, 2017, pp. 2353-2356.
- [51] Juan Quiroz, Elena Geangu, and Min Hooi Yong, "Emotion-Recognition Using Smart Watch Sensor Data: Mixed-Design Study," *JMIR Mental Health*, vol. 5, May 2018.
- [52] (2021, Octubre) Recurrent_neural_network. [Online]. https://en.wikipedia.org/wiki/Recurrent_neural_network
- [53] Giuseppe Riva et al., "Affective Interactions Using Virtual Reality: The Link between Presence and Emotions," *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, vol. 10, pp. 45-56, Mar. 2007.
- [54] David Rubin and Jennifer Talarico, "A Comparison of Dimensional Models of Emotion: Evidence from Emotions, Prototypical Events, Autobiographical Memories, and Words," *Memory (Hove, England)*, vol. 17, pp. 802-8, Sep. 2009.
- [55] James Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161-1178, Dec. 1980.
- [56] Stanisław Saganowski et al., Emotion Recognition Using Wearables: A Systematic Literature Review - Work-in-progress, Mar. 2020.

- [57] Luz Santamaria-Granados, Mario Organero, Gustavo Ramirez-Gonzalez, Enas Abdulhay, and Arunkumar N., "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)," *IEEE Access*, vol. PP, pp. 1-1, Nov. 2018.
- [58] Yeong-Seok Seo and Jun-Ho Huh, "Automatic Emotion-Based Music Classification for Supporting Intelligent IoT Applications," *Electronics*, vol. 8, p. 164, Feb. 2019.
- [59] Karan Sharma, Claudio Castellini, Egon L. van den Broek, Alin Albu-Schäeffler, and Friedhelm Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," Dec. 2018.
- [60] Shu et al., "Wearable Emotion Recognition Using Heart Rate Data from a Smart Bracelet," *Sensors*, vol. 20, p. 718, Jan. 2020.
- [61] Lin Shu et al., "A Review of Emotion Recognition Using Physiological Signals," *Sensors*, vol. 18, p. 2074, June 2018.
- [62] Ewa Siedlecka and Thomas Denson, "Experimental Methods for Inducing Basic Emotions: A Qualitative Review," *Emotion Review*, vol. 11, p. 175407391774901, Mar. 2018.
- [63] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic, "A Multi-Modal Affective Database for Affect Recognition and Implicit Tagging," *Affective Computing, IEEE Transactions on*, vol. 3, p. 1, Jan. 2012.
- [64] Madeena Sultana, Majed Al-Jefri, and Joon Lee, "Using Machine Learning and Smartphone and Smartwatch Data to Detect Emotional States and Transitions: An Exploratory Study," *JMIR mHealth and uHealth*, vol. 8, Jan. 2020.
- [65] Jianhua Tao and Tieniu Tan, "Affective Computing: A Review," Oct. 2005.
- [66] Alexander Toet, Fabienne Heijn, Anne-Marie Brouwer, Tina Mioch, and Jan Erp, "The EmojiGrid as an Immersive Self-report Tool for the Affective Assessment of 360 VR Videos," in *Virtual Reality and Augmented Reality.*, 2019, pp. 330-335, isbn [978-3-030-31907-6].
- [67] Alexander Toet, Fabienne Heijn, Anne-Marie Brouwer, Tina Mioch, and Jan B. F. van Erp, "An Immersive Self-Report Tool for the Affective Appraisal of 360° VR Videos," *Frontiers in Virtual Reality*, vol. 1, Sep. 2020.
- [68] (2021, Oct.) Understanding LSTM Networks -- colah's blog. [Online]. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [69] Ielka Van der sluis, Chris Mellish, and Gavin Doherty, "Affective Text: Generation Strategies and Emotion Measurement Issues," in *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference, May 18-20, 2011, Palm Beach, Florida, USA*, 2011.
- [70] Pauli Virtanen et al., "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, pp. 1-12, Feb. 2020.

- [71] Jonathan Vitale, Mary-Anne Williams, Benjamin Johnston, and Giuseppe Boccignone, "Affective facial expression processing via simulation: A probabilistic model," *BIOLOGICALLY INSPIRED COGNITIVE ARCHITECTURES*, vol. 10, pp. 30-41, Oct. 2014.
- [72] (2021, Octubre) Watson and Tellegen's (1985) Circumplex Model. *Psychology* 425. [Online]. <https://psu.pb.unizin.org/psych425/chapter/watson-and-tellegens-1985-circumplex-model/>
- [73] (2021, Octubre) WHAT IS THE MAIN DIFFERENCE BETWEEN RNN AND LSTM. [Online]. <https://ashutoshtripathi.com/2021/07/02/what-is-the-main-difference-between-rnn-and-lstm-nlp-rnn-vs-lstm/>
- [74] (2021, Octubre) what-is-a-neural-network. [Online]. <https://www.tibco.com/reference-center/what-is-a-neural-network>
- [75] Simon Whatley. (2013, Abril) Plutchik's wheel of emotion. [Online]. <https://www.simonwhatley.co.uk/writing/plutchik-wheel-of-emotion/>
- [76] Dongrui Wu et al., "Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task," *IEEE Transactions on Affective Computing*, vol. 1, pp. 109-118, July 2010.
- [77] Chih-Hung Wu, Yueh-Min Huang, and Jan-Pan Hwang, "Review of affective computing in education/learning: Trends and challenges," *British Journal of Educational Technology*, vol. 47, Sep. 2015.
- [78] Tianhua Xie, Mingliang Cao, and Zhigeng Pan, "Applying Self-Assessment Manikin (SAM) to Evaluate the Affective Arousal Effects of VR Games," in *ICIGP 2020: 2020 3rd International Conference on Image and Graphics Processing*, Feb. 2020, pp. 134-138.
- [79] Xiaofen Xing et al., "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Frontiers in Neurorobotics*, vol. 13, June 2019.
- [81] Joseph Yacim and Douw Boshoff, "Impact of Artificial Neural Networks Training Algorithms on Accurate Prediction of Property Values," *Journal of Real Estate Research*, vol. 40, pp. 375-418, Nov. 2018.
- [82] Shen Zhang, Zhiyong Wu, Helen Meng, and Lianhong Cai, "Facial Expression Synthesis Based on Emotion Dimensions for Affective Talking Avatar," *Smart Innovation, Systems and Technologies*, vol. 2010, pp. 109-132, June 2010.