

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ingeniería



Informe Final Trabajo Profesional de Ingeniería Informática

Herramienta para la evaluación de emociones en contextos abiertos

Integrantes

- Santiago Fernandez sfernandezc@fi.uba.ar
- Maria Sol Fontenla msfontenla@fi.uba.ar
- Ignacio Iragui iiragui@fi.uba.ar
- Agustina Segura asegura@fi.uba.ar

Grupo: 21

Tutor: Dr. Jorge Ierache, jierache@fi.uba.ar

Índice de contenidos

Resumen.....	5
Abstract.....	6
Palabras clave.....	7
Keywords.....	7
Agradecimientos.....	8
1. Introducción.....	9
2. Estado del Arte.....	11
2.1. Expresiones Faciales Universales.....	11
2.1.1 Sistema de Codificación Facial (FACS).....	11
2.2. Computación Afetiva.....	14
2.2.2 Enfoque categórico.....	15
2.2.3 Enfoque dimensional.....	16
2.2.4 Reconocimiento de emociones en imágenes y videos.....	18
2.3 Estímulos y Encuesta SAM.....	18
2.4. Extracción de características faciales y emociones.....	21
2.4.1 Servicios para detección de LandMarks.....	22
2.4.1.1 OpenFace.....	22
2.4.1.2 PyFeat.....	23
2.4.1.3 Face ++.....	24
2.4.2 Servicios para detección de emociones.....	24
2.4.2.1 Face ++.....	25
2.4.2.2 ONNX FER+ Emotion Recognition.....	25
2.4.2.3 PyFeat.....	26
2.4.2.4 AWS Rekognition.....	27
3. Problema detectado y/o faltante.....	28
4. Solución implementada.....	30
4.1 Introducción.....	30
4.2 Detalle de los componentes.....	31
4.2.1 API.....	31
4.2.2 Procesadores.....	33
4.2.2.1 Valence processor.....	33
4.2.2.2 Arousal processor.....	35
4.2.3 Joiner.....	39
4.2.4 Aplicación Web.....	40
4.2.4.1 Login.....	40
4.2.4.2 Registro.....	41
4.2.4.3 Nuevo Análisis Video.....	42
4.2.4.4 Nuevo Análisis Imagen.....	42
4.2.4.5 Mis Videos.....	43

4.2.4.6 Mis Imágenes.....	43
4.2.4.7 Sección Modelo Ekman y Modelo Russel.....	44
4.2.4.8 Sección carga de video a procesar y video estímulo.....	45
4.2.4.9 Sección Carga de imagen a procesar y estímulo.....	46
4.2.4.10 Sección unidades de acción.....	47
4.2.4.11 Sección Excitación y Valencia en el Tiempo.....	48
4.2.4.12 Sección Resumen.....	49
4.3 Arquitectura y Especificaciones técnicas.....	50
4.3.1 Registro y login de usuarios.....	50
4.3.2 Comunicación Joiner-API.....	50
4.3.2 Persistencia de los datos.....	51
4.3.3 Comunicación Web-API.....	53
4.3.4 Procesamiento de Frames.....	53
4.3.4 Repositorios adicionales.....	55
5. Metodología aplicada.....	56
5.1 Organización del Trabajo.....	56
5.2 Herramienta de Trabajo.....	56
6. Experimentación y/o validación.....	57
6.1 Pruebas basales.....	57
6.1.1 Asco.....	57
6.1.2 Miedo.....	58
6.1.3 Felicidad.....	60
6.1.4 Tristeza.....	61
6.1.5 Sorpresa.....	62
6.1.6 Enojo.....	63
6.2 Pruebas de excitación y valencia comparadas con dataset DEVO.....	65
6.2.1 Video estímulo 28.1.....	66
6.2.1.1 Nazareno.....	67
6.2.1.2 Valentina.....	68
6.2.1.3 Deborah.....	70
6.2.1.4 Conclusión para el video 28.1.....	71
6.2.2 Video 96.1.....	71
6.2.2.1 Deborah.....	72
6.2.2.2 Conclusion Video 96.1.....	73
6.2.3 Video 21.6.....	73
6.2.3.1 Roma.....	74
6.2.3.2 Valentina.....	76
6.2.3.3 Conclusion Video 21.6.....	77
6.2.4 Pruebas de valencia y excitación - conclusiones.....	77
6.3 Prueba de campo.....	78
6.4 Validación con Morphcast.....	81
6.4.1 Análisis para prueba de campo.....	81
6.4.2 Análisis de reacciones a videos de DEVO.....	86
6.4.2.1 Video 28.1 - Valentina.....	86

6.4.2.2 Video 21.6 - Roma.....	87
6.4.2.3 Video 96.1 - Deborah.....	89
6.4.3 Validación con MorphCast - conclusiones.....	90
7. Cronograma de las actividades realizadas.....	91
7.1 Hitos de avance.....	91
7.2 Matriz de tiempo de los Hitos.....	92
7.3 Tareas realizadas por mes.....	92
8. Riesgos materializados y lecciones aprendidas.....	93
8.1 Riesgos.....	93
8.2 Lecciones Aprendidas.....	94
9. Trabajos futuros.....	95
10. Conclusiones.....	97
Referencias.....	99
Anexos.....	103
Manual de ejecución.....	103
Código fuente del sistema.....	103
Comparación de frames a procesar.....	103
Comparativa FerPlus - PyFeat.....	103
Comparativa unidades de acción.....	103
Pruebas de excitación y valencia.....	103
Pruebas de Campo.....	103
Tareas realizadas por mes.....	104

Resumen

Las emociones desempeñan un papel fundamental en la comunicación, éstas agregan capas de significado y enriquecen nuestra capacidad de conexión y comprensión en la comunicación cotidiana. Nuestros gestos, expresiones faciales, tono de voz y postura corporal ayudan a transmitir una gran parte de un mensaje. Esta comunicación no verbal, a través de señales sutiles o expresiones faciales, permite transmitir emociones como alegría, tristeza, enojo o empatía. En la última década, gracias a los avances tecnológicos de procesamientos de imágenes y videos se han desarrollado numerosas investigaciones acerca del reconocimiento automático de estas emociones a partir de las expresiones faciales.

En la actualidad existen numerosos modelos y desarrollos de software que permiten reconocer automáticamente las emociones, en el presente trabajo, se propone explorar los modelos categóricos y dimensionales combinándolos para la detección de emociones. El sistema cuya implementación se detalla en la solución propuesta permite clasificar seis emociones básicas y otras 32 emociones a partir del cálculo de la valencia (valence) y la excitación (arousal) utilizando las expresiones faciales obtenidas de los rostros de las personas al utilizar el sistema en base a un contexto dado.

Finalmente, se detallan distintas pruebas individuales e integrales que se llevaron a cabo con el fin de ampliar el espectro de conclusiones obtenidas. Se presentarán los resultados obtenidos en conjunto con un análisis de los mismos.

Abstract

Emotions play a fundamental role in communication, they add layers of meaning and enrich our ability to connect and understand in everyday communication. Our gestures, facial expressions, tone of voice, and body posture help convey a significant part of a message. This nonverbal communication, through subtle signals or facial expressions, allows the transmission of emotions such as joy, sadness, anger, or empathy. In the last decade, numerous research studies have been developed thanks to technological advances in image and video processing, focusing on the automatic recognition of these emotions from facial expressions.

Currently, there are numerous models and software developments that enable the automatic recognition of emotions. In this work, the proposal explores categorical and dimensional models, combining them for emotion detection. The system detailed in the proposed solution classifies six basic emotions and other 32 emotions based on the calculation of valence and arousal using facial expressions obtained from individuals when using the system based on a given context..

Finally, several individual and comprehensive tests were detailed and carried out to expand the spectrum of conclusions obtained. The results are presented alongside an analysis of the findings.

Palabras clave

- Computación afectiva, expresiones faciales, emociones básicas, valencia, excitación, unidades de acción, puntos de referencia, estímulos

Keywords

- Affective computing, facial expressions, basic emotions, valence, arousal, action units, landmarks, stimulus

Agradecimientos

Queremos expresar nuestro más sincero agradecimiento a todas las personas que hicieron posible la realización de este trabajo final.

En primer lugar, agradecemos a nuestro tutor, el Dr. Jorge Ierache, por su valiosa orientación, apoyo constante y paciencia a lo largo de todo el proceso de desarrollo. Su experiencia y conocimientos fueron fundamentales para el desarrollo del mismo.

Agradecemos también a la Universidad de Buenos Aires por brindarnos los recursos necesarios y un ambiente de aprendizaje constante.

Un reconocimiento especial para nuestras familias y amigos, por su paciencia, comprensión y apoyo incondicional durante todo el tiempo que dedicamos a este proyecto y a nuestra carrera. Sin su apoyo emocional, esta tarea hubiera sido mucho más difícil de llevar a cabo.

Finalmente, agradecemos a todas las personas y entidades que, de una forma u otra, aportaron su granito de arena para que este proyecto se hiciera realidad.

1. Introducción

Sabemos que las emociones juegan un rol muy importante en la vida cotidiana de las personas. Estas influyen en todos los aspectos de nuestra vida, principalmente en nuestra manera de relacionarnos con los demás y en cómo percibimos el mundo.

En los últimos años la manera de vincularnos se vio impactada por los avances tecnológicos y donde más se vio este impacto fue en la pandemia. Desde 2020 se acrecentaron los encuentros vía plataformas virtuales, principalmente por motivos sanitarios y ya estando en 2024 vemos que esta forma de comunicarnos, encontrarnos y relacionarnos se instaló para quedarse. Como aspecto positivo, esto permitió que nos podamos conectar de una manera más fluida, al alcance de nuestro hogar y desde cualquier parte del mundo. Sin embargo, también trajo como consecuencia una alteración en la percepción de las emociones, ya que a través de una pantalla es más difícil percibir el estado emocional de una persona, e incluso el estado emocional colectivo de todos aquellos involucrados en los encuentros.

Dentro del ámbito tecnológico existen numerosos estudios sobre cómo capturar las emociones de los seres humanos, enfocados en poder interpretar las emociones de los usuarios por medio de una computadora a través de un video o imagen. Particular el enfoque está dado sobre analizar las expresiones faciales para poder revelar el estado emocional de una persona.

Para este trabajo, planteamos como objetivo la implementación de un sistema capaz de reconocer una serie de emociones a partir de una imagen o video dentro de un contexto, empleando una combinación entre el modelo categórico y el modelo dimensional. A su vez, este sistema proporciona un análisis integral al permitir incorporar el estímulo al cual dicha persona estaba siendo expuesta. El trabajo surge a partir de que hoy en día no hay sistemas gratuitos o de código libre que proporcionen un análisis en profundidad sobre la detección de emociones propuestas tanto por el modelo categórico como el modelo dimensional en conjunto con el estímulo.

En la sección 2, se presenta el estudio del estado del arte del reconocimiento de emociones de personas en videos e imágenes a través del análisis de expresiones faciales, utilizando distintos modelos. Se menciona la importancia que juegan las expresiones faciales al momento de determinar las emociones. Se describen dos modelos junto a su procedimiento de reconocimiento automático de emociones en imágenes y videos y la importancia de los estímulos al momento de realizar un análisis emocional completo.

En la sección 3, se definen los problemas encontrados hoy en día respecto a la falta de sistemas de detección emocional en videos e imágenes. La falta de sistemas eficientes y accesibles para el reconocimiento emocional es uno de los desafíos principales. Se plantea una solución potencial, detallando cómo nuestra propuesta puede superar estas limitaciones

Herramienta para la evaluación de emociones en contextos abiertos

mediante el uso de una combinación de modelos categóricos y dimensionales, junto con la incorporación de estímulos contextuales.

En la sección 4, se presenta la solución propuesta en detalle. Se describe la arquitectura del sistema, incluyendo sus componentes principales y la interacción entre ellos. Se explica cómo se integran los diferentes modelos y tecnologías para lograr un reconocimiento emocional preciso y confiable. Esta sección proporciona una visión clara de la estructura y funcionamiento del sistema propuesto.

En la sección 5, se detalla la metodología aplicada detallando el enfoque utilizado, la organización del trabajo y la gestión del proyecto.

En la sección 6, se presentan los resultados de las pruebas realizadas al sistema. Se describen las pruebas individuales e integrales donde se pone a prueba el sistema en distintos contextos. Esta sección proporciona una evaluación crítica del sistema, basándose en datos empíricos y análisis detallados.

En la sección 7, se describe el cronograma seguido por el equipo de trabajo. Se presentan las fases del proyecto, los hitos importantes y los plazos establecidos. Esta sección ofrece una visión del proceso de planificación y ejecución del proyecto.

En la sección 8, se analizan los riesgos enfrentados durante el desarrollo del proyecto y las lecciones aprendidas. Se describen los problemas y desafíos que surgieron, cómo fueron abordados y las estrategias implementadas para mitigar los riesgos.

Finalmente en la sección 9 se detallan futuros desarrollos para la continuación del proyecto. Se proponen mejoras, nuevas áreas de investigación y aplicación y extensiones que podrían implementarse para aumentar la funcionalidad y eficiencia del sistema. Esta sección ofrece una visión a largo plazo del proyecto y sus posibles evoluciones.

2. Estado del Arte

La sección de "Estado del Arte" proporciona una revisión crítica de la literatura existente en el campo de la computación afectiva, con un enfoque particular en las expresiones faciales universales, los modelos categóricos y dimensionales para el reconocimiento de las emociones, los estímulos y encuestas SAM (Self-Assessment Manikin) [41], y la extracción de características faciales.

2.1. Expresiones Faciales Universales

Todos los seres humanos conocemos el concepto de emoción e incluso lo utilizamos para definir en qué estado nos encontramos en nuestro día a día. Klaus R. Scherer, profesor y director en Swiss Centre for Affective Sciences (Geneva, Francia) en el año 2005 [1], define una emoción como '*un episodio de cambios interrelacionados y sincronizados ... en respuesta de la evaluación de un evento externo o interno...*'. Además, expresa que para medir las emociones se necesita una serie de procesos involucrados en las emociones, pero que han habido avances en la medición de componentes individuales de las emociones. Uno de dichos componentes son las expresiones faciales y estas pueden medirse para *inferir* la emoción.

Habiendo enunciado esto, Paul Ekman [2], psicólogo estadounidense pionero en el estudio de las emociones y su expresión facial, describió en 1984 seis emociones universales básicas: felicidad, miedo, tristeza, enojo, asco y sorpresa (en inglés “happiness”, “fear”, “sadness”, “anger”, “disgust”, “surprise”), sumando una séptima en su propuesta: desprecio (en inglés “contempt”). [3] [4].

Otro avance significativo en la investigación de Paul Ekman [2] ocurrió en 1978, cuando, junto a su colega Walter Friesen, publicaron el Sistema de Codificación de la Actividad Facial (FACS) [8]. Esta herramienta les permitió caracterizar en detalle cada una de las siete expresiones faciales mencionadas anteriormente.

Gracias a los estudios de Ekman es que hoy se pueden inferir las emociones de un individuo mediante modelos de aprendizaje automático, tanto discretos como dimensionales, que enunciamos a continuación.

2.1.1 Sistema de Codificación Facial (FACS)

El Sistema de Codificación de la Actividad Facial (FACS) [8] es una herramienta que permite identificar y categorizar los movimientos faciales. El objetivo principal del FACS es desglosar las expresiones faciales en sus componentes más básicos. Cada expresión se analiza en términos de unidades de acción (AUs, por sus siglas en inglés), que representan las contracciones individuales de un músculo o grupo de músculos específicos. Esta

Herramienta para la evaluación de emociones en contextos abiertos

descomposición permite una descripción precisa y detallada de cada movimiento facial, facilitando así su estudio y comprensión.

Este sistema define un total de 46 unidades de acción donde cada una de ellas define un determinado cambio en el rostro provocado por la acción de diferentes músculos. En la figura 1 se muestran las unidades de acción utilizadas para la detección de emociones.

Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28

Figura 1: Unidades de acción para la parte superior e inferior del rostro

Fuente: J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the Facial Action Coding System, " The handbook of emotion elicitation and assessment, Oxford University Press Series in Affective Science, New York: Oxford, 2007.

A su vez, como se puede observar en la figura 2, este sistema le otorga a cada AU un nivel de intensidad basado en una escala de 5 niveles. Los niveles definidos son A, B, C, D y E; siendo A ("Trace") el de menor intensidad y E ("Maximum") el de mayor intensidad. Es importante mencionar que estos niveles de intensidad no poseen el mismo intervalo.

Herramienta para la evaluación de emociones en contextos abiertos

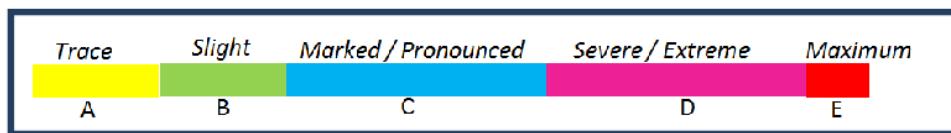


Figura 2: Escala de niveles de intensidad para las AUs.

Fuente: O. Rudovic, V. Pavlovic and M. Pantic, "Context-Sensitive Dynamic Ordinal Regression for Intensity Estimation of Facial Action Units" in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 37, no. 05, pp. 944-958, 2015.

El FACS permite vincular estas unidades de acción con cada una de las expresiones que se pretende detectar, es decir, permite determinar qué unidades de acción se encuentran en cada una de las emociones básicas propuestas por Paul Ekman [2]. Esto lo podemos ver en la siguiente figura.

Emotion	AUs
Anger	4+5+7+10+22+23+25/26
	4+5+7+10+23+25/26
	4+5+7+17+23/24
	4+5+7+23/24
	4+5/7
	17+24
Disgust	9/10+17
	9/10+16+25/26
	9/10
Fear	1+2+4
	1+2+4+5+20+25/26/27
	1+2+4+5+25/26/27
	1+2+4+5
	1+2+5+25/26/27
	5+20+25/26/27
	5+20
	20
	12
Happy	6+12
	1+4
	1+4+11/15
	1+4+15+17
Sadness	6+15
	11+17
	1
	1+2+5+26/27
	1+2+5
Surprise	1+2+26/27
	5+26/27

Figura 3: Reglas de mapeo entre unidades de acción y emociones.

Fuente: The Facial Action Coding System for Characterization of Human AffectiveResponse to Consumer Product-Based Stimuli: A Systematic Review. Elizabeth A. Clark, J’Nai Kessinger, Susan E. Duncan, Martha Ann Bell, Jacob Lahne, Daniel L. Gallagher and Sean F. O’Keefe. 2020.

2.2. Computación Afectiva

La computación afectiva es una rama de la informática que busca poder a partir de distintas expresiones digitales poder reconocer, interpretar, procesar y estimular el estado emocional de un usuario, entendiendo a éste desde un análisis categórico o discreto. [5]

La mayor dificultad presentada para poder avanzar en cuestiones de computación afectiva se presenta en la idea humana preconcebida de que las emociones son una propiedad puramente humana y no que se presentan como una herramienta utilizada para facilitar la comunicación entre dos entes, sean humanos o no. [6]

Entender a las emociones como una herramienta utilizada a la hora de facilitar la comunicación nos permite extender nuestro estudio de emociones a un sinfín de expresiones del usuario siempre y cuando correspondan a un medio comunicativo. Este es uno de los motivos por los que hoy en dia es un área de estudio en gran crecimiento ya que en una sociedad globalizada como la de hoy en día, la comunicación está presente en todos los ámbitos y de allí se pueden extraer un estudio de las emociones, que provee una mirada nueva a cualquier enfoque previo.

Como primer enfoque dentro del área, se encuentra el reconocimiento de emociones a partir de expresiones del usuario. Dentro de las distintas áreas que se utilizan para reconocer el estado emocional se destacan:

- Imagen (Fotografía y video)
- Texto
- Voz
- Respuestas Fisiológicas (EEG, variación de ritmo cardiaco, conductancia de piel)
- Biométricas

Una vez reconocida la emoción se presenta toda una etapa de interpretación y procesamiento de la misma que eso dependerá del contexto en el que se encuentre y el enfoque que se tome (discreto o categórico). Por último se presenta el desafío de poder comprender la interpretación para llevarla a un sistema informático que emule estas emociones para poder estimular al usuario.

Se entiende también que no todo proyecto de computación afectiva debe abarcar todas las fuentes ni todos los análisis mencionados anteriormente. Lo esencial es que estén directamente relacionados con la comprensión del estado emocional del usuario y se adapten a un contexto específico.

2.2.2 Enfoque categórico

El modelo categórico propuesto por el psicólogo Paul Ekman [7][8] establece la teoría de que hay un conjunto básico y universal de las emociones que pueden ser expresadas por medio de expresiones faciales.

Inicialmente se plantea que existen 6 emociones que son reconocidas y aceptadas en diversas culturas: felicidad, tristeza, enojo, sorpresa, asco y miedo. Más adelante, Ekman propone incorporar una séptima expresión facial que representa la emoción del “desprecio”. En la Figura 4 se pueden observar las distintas emociones representadas por actores.

Estas emociones son consideradas universales debido a que se manifiestan de forma similar en las expresiones faciales de todas las culturas, independientemente de las diferencias culturales y lingüísticas. Estas emociones no dependen ni del idioma, región, cultura o etnia.

El modelo se basa en la observación y análisis detallado de las expresiones faciales asociadas con cada emoción. Ekman desarrolló el sistema de codificación de acciones faciales (FACS) cuyo propósito es detectar los cambios que se producen en el rostro debido a los músculos faciales y así poder caracterizar las expresiones faciales [8][9]. A estas acciones se las denomina AUs (Unidades de Acción por su sigla en inglés) y son movimientos musculares específicos que se producen en la cara y están relacionados con una emoción particular.



Figura 4: 7 expresiones faciales propuesta por Ekman

Fuente: Ekman, P., and Friesen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial

Cada emoción básica se caracteriza por una combinación única y reconocible de unidades de acción facial. Estas combinaciones de unidades de acción se pueden identificar y cuantificar utilizando el FACS, permitiendo una comprensión precisa y estandarizada de las expresiones faciales.

2.2.3 Enfoque dimensional

En el año 1980, el psicólogo James A. Russell definió que las emociones pueden describirse como un set de dimensiones [10][11] en su modelo circumplejo. En particular, dos dimensiones a las cuales describe como valencia y excitación, y categoriza a las emociones en términos de placer y grado de arousal.

La valencia indica si el estado emocional de la persona es positivo o negativo. "Feliz" es la única expresión positiva, "triste", "enojado", "asustado" y "disgustado" se consideran expresiones negativas.

Para calcular la valencia este modelo utiliza las categorías propuestas por el psicólogo Paul Ekman [4] y su intensidad asociada. La valencia se calcula como la intensidad de "feliz" menos la intensidad de la expresión negativa con la intensidad más alta. Por ejemplo, si la intensidad de "feliz" es 0.8 y la intensidad de "triste", "enojado", "asustado" y "Disgustado" son 0.2, 0.0, 0.3 y 0.2, respectivamente, entonces la valencia es $0.8 - 0.3 = 0.5$.

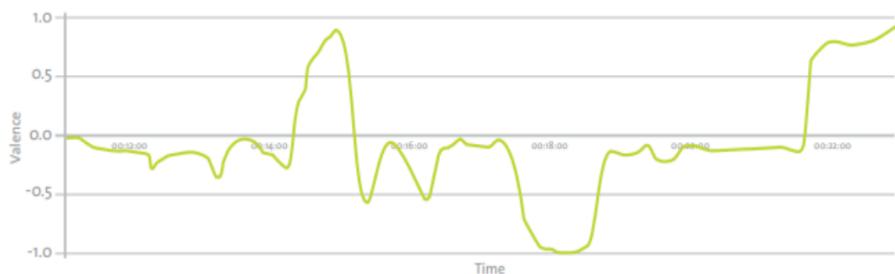


Figura 5: Variación de valencia respecto al tiempo.

Fuente: O. Krips and L. Loijens, “FaceReader Methodology Note,” A White Pap. by Noldus Inf. Technol.

Por otro lado, el arousal es la capacidad de estar despierto y de mantener la alerta que implica la capacidad de seguir estímulos u órdenes, está basado en la activación de 20 Unidades de Acción (AU) del Sistema de Codificación de Acción Facial (FACS). Para poder calcularla se requiere:

1. Calcular las unidades de acción (AU) y sus valores de activación (AV)

Herramienta para la evaluación de emociones en contextos abiertos

2. Utilizar los valores de activación (1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 24, 25, 26, 27 y 43) y calcula el promedio de estos en los últimos 60 segundos de acuerdo con la expresión

$$AAV_j = \frac{1}{N} \sum_{i=N}^{i=1} (AV_j)$$

donde n es igual a 60 que implica los últimos 60 segundos pasados que se toma para el cálculo del promedio de la j-ésima unidad de acción.

3. Se calcula la diferencia entre el AAV Actual y el AAV promedio. Esto se hace con el fin de corregir aquellas unidades de acción que se encuentran continuamente activas y podrían indicar un sesgo individual. Esto da como resultado el valor corregido del valor de activación (CAV o “Corrected Activation Value”)

4. La excitación se calcula como el valor promedio entre los últimos 5 CAV

Finalmente se utilizan los valores de valencia (eje horizontal visible en la figura 6) y del arousal (eje vertical) para ser ubicados cada uno en su correspondiente eje y así poder detectar qué emociones tuvieron una mayor presencia. [22]

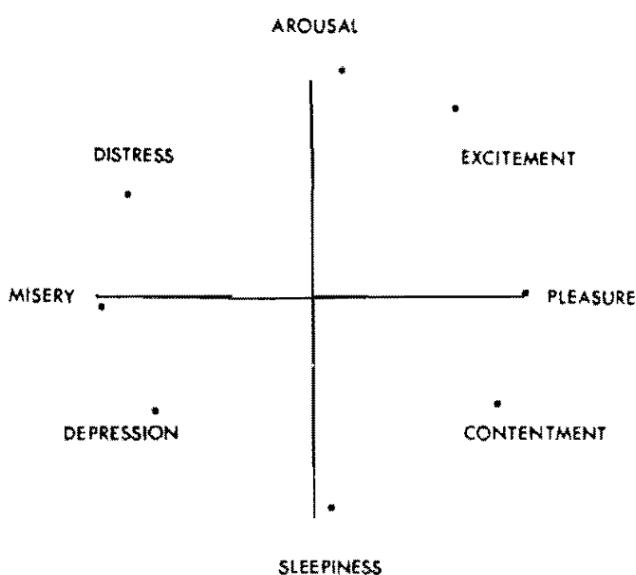


Figura 6: modelo dimensional de Russell.

Fuente: Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178.

El modelo dimensional divide el mapa en 8 categorías en orden circular: excitación, contento, deprimido, angustiado, miserable, complacido, activado y somnoliento (en inglés “excitement”, “contentment”, “depressed”, “distressed”, “misery”, “pleasure”, “arousal” y “sleepiness”). [12]

2.2.4 Reconocimiento de emociones en imágenes y videos

El reconocimiento de emociones en imágenes es un proceso que busca capturar e interpretar las emociones expresadas en los rostros de las personas en una imagen. Este proceso se basa en el concepto de que las emociones humanas se manifiestan de manera característica en las expresiones faciales.

Una técnica para poder reconocer las emociones en las imágenes es por medio de algoritmos basados en inteligencia artificial, particularmente en el campo del aprendizaje profundo. Estos algoritmos son entrenados a partir de distintos sets de datos, entre ellos podemos encontrar FER+[16], eLFW[20] CK+[21] en cual luego predicen la emoción predominante en la imagen o una serie de emociones predominante junto a su porcentaje de probabilidad. Estos modelos generalmente utilizan las emociones categorías planteadas por el psicólogo Paul Ekman [4] y agregan al set de emociones la emoción “neutra”.[23][24]

Además de los modelos categóricos mencionados, el reconocimiento de emociones también puede abordarse desde una perspectiva dimensional. En este enfoque, las emociones se representan en un espacio continuo definido por dimensiones como la valencia y la excitación. Para implementar estos modelos dimensionales se requiere de la utilización de landmarks y unidades de acción que se detallarán en la siguiente sección.

2.3 Estímulos y Encuesta SAM

El análisis de las emociones detectadas en imágenes y videos está intrínsecamente relacionado a los estímulos que se presentan a los individuos. Un estímulo puede ser cualquier elemento, evento o condición que haga que un organismo, como una persona, responda o reaccione. Los estímulos pueden incluir situaciones, escenarios o interacciones sociales que provocan respuestas emocionales, cognitivas o conductuales. Estos pueden ser:

- Imágenes
- VideoJuegos
- Videos
- Música / Sonidos
- Olores

Para nuestro estudio nos enfocaremos en estímulos de imágenes y videos que van a desempeñar un papel crucial en el estudio de las respuestas emocionales humanas.

La importancia de los estímulos radica en su capacidad para evocar respuestas emocionales específicas, las cuales pueden ser evaluadas de manera objetiva. Al presentar un estímulo visual, como una imagen o un video, se puede observar y medir la reacción emocional de una persona mediante herramientas como la Escala de Respuestas Afectivas Autoinformadas (SAM) [41]. Esta encuesta permite a los participantes calificar su experiencia emocional en términos de valencia (si la experiencia es positiva o negativa) y excitación (el nivel de activación o arousal). En la figura 7 se puede observar un ejemplo de encuesta sam donde el usuario puede seleccionar tanto para la valencia como la excitación un valor del 1 al 9.

Encuesta SAM

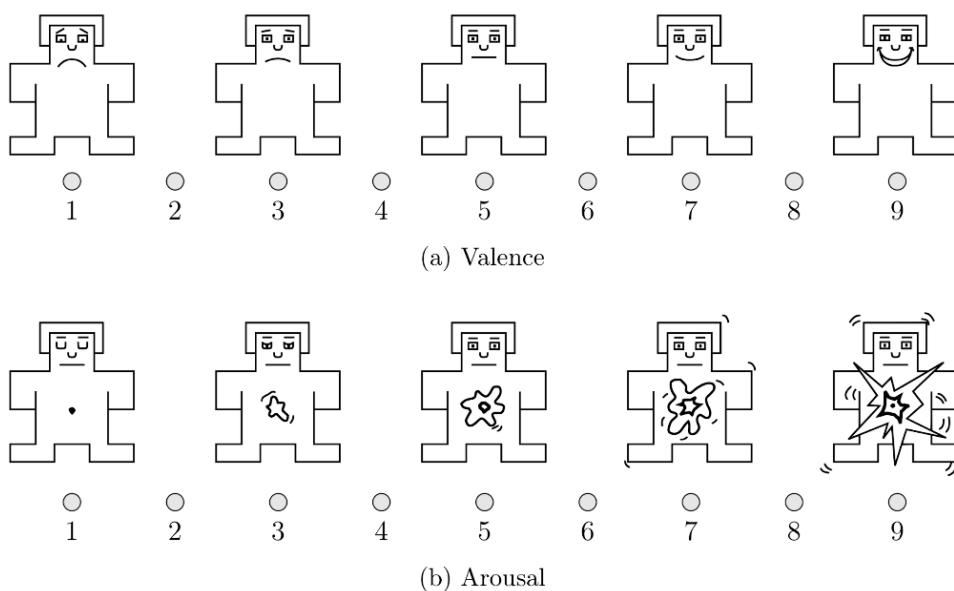


Figura 7: Ejemplo de encuesta SAM asociado a Excitación y Valencia.

Fuente: Ierache, Jorge & Sattolo, Iris & Chapperón Gabriela,. (2021). Framework multimodal emocional en el contexto de ambientes dinámicos. RISTI - Revista Ibérica de Sistemas y Tecnologías de Informacao. 45-59. 10.17013/risti.40.45-59.

La relación entre los estímulos y la encuesta SAM [41] es crucial para obtener datos consistentes y comparables. Si una persona reacciona a un estímulo y luego se le aplica la encuesta SAM, se puede determinar la valencia y excitación asociadas a dicho estímulo. Este proceso estandarizado permite que, al presentar el mismo estímulo a diferentes personas, se esperen respuestas emocionales similares en términos de valencia y excitación promedio, lo que facilita la evaluación y comparación de las reacciones emocionales.

Herramienta para la evaluación de emociones en contextos abiertos

Para el presente trabajo, se utilizan los datasets DEVO [42] e IAPS [43] como fuentes de estímulos para realizar las pruebas. Estos datasets han sido ampliamente utilizados en investigaciones psicológicas y de computación afectiva debido a su capacidad para evocar respuestas emocionales consistentes y medibles.

El dataset DEVO [42] (Database of Emotional Videos) contiene una serie de videos diseñados para evocar emociones específicas. Cada video en este dataset ha sido previamente evaluado y catalogado en términos de su valencia y excitación promedio, proporcionando un marco de referencia claro para el análisis de las reacciones emocionales de los participantes.

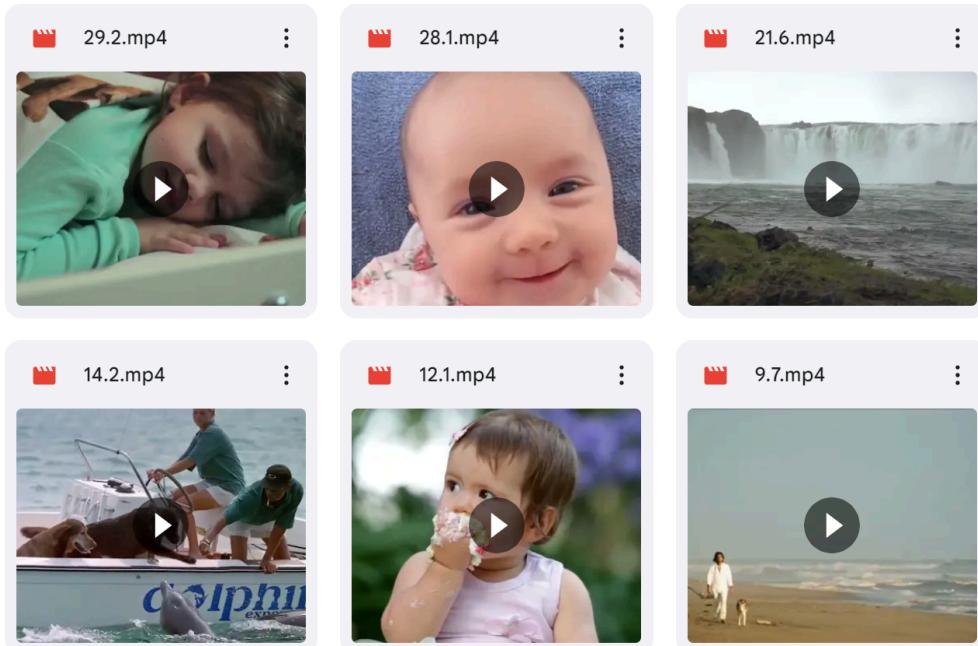


Figura 8: Ejemplo de videos pertenecientes al dataset DEVO.

Fuente: Ack Baraly, K. T., et al. (2020). Database of Emotional Videos from Ottawa (DEVO).

Collabro: Psychology, 6(1): 10. DOI: <https://doi.org/10.1525/collabra.180>.

En la figura 8 se puede observar una serie de ejemplos de videos pertenecientes al set de datos de DEVO.

El IAPS [43] (International Affective Picture System) es un conjunto de imágenes estandarizadas que también ha sido evaluado en términos de valencia y excitación. Este dataset es ampliamente utilizado en estudios de emociones debido a su robustez y la fiabilidad de las respuestas emocionales que evoca. En la figura 9 se puede observar una serie de imágenes de distintos estímulos correspondientes al set de datos de IAPS..



Figura 9: Ejemplo de imágenes pertenecientes al dataset de IAPS.

Fuente: Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. Informe Técnico A-8. University of Florida, Gainesville, FL.

2.4. Extracción de características faciales y emociones

La extracción de características faciales es un proceso fundamental en el análisis de imágenes y videos que consiste en obtener información significativa del rostro de una persona. Este proceso es crucial tanto para el entrenamiento de modelos predictivos como para la realización de predicciones a partir de modelos ya entrenados. De entre todas las características que podemos obtener las que más nos interesan son un conjunto de puntos o **landmarks** faciales que permiten determinar regiones del rostro (contorno de la cara, cejas, ojos, etc.) o bien las emociones detectadas a partir de una imagen de un rostro. Esta tarea se realiza para obtener la entrada necesaria para el entrenamiento de un modelo predictivo o para realizar predicciones a partir de un modelo ya entrenado.

La tarea de detectar landmarks es esencial para identificar las unidades de acción, que son utilizadas posteriormente para asignar emociones. Esta técnica permite abstraer la información específica de la foto a un modelo completamente matemático, eliminando las particularidades individuales de cada imagen. De esta manera, se obtiene la entrada necesaria para el entrenamiento de un modelo predictivo o para realizar predicciones con un modelo ya entrenado.

Existe una gran variedad de API's y modelos pre-entrenados disponibles para realizar estas tareas y a continuación presentamos diferentes librerías y servicios que fueron relevados para el presente trabajo.

2.4.1 Servicios para detección de LandMarks

A continuación se detallan servicios y herramientas para poder detectar Landmarks faciales en imágenes y/o vídeos.

2.4.1.1 OpenFace

OpenFace es un proyecto de código abierto que brinda un kit de herramientas capaz de detectar puntos de referencia faciales (landmarks), estimar la posición de la cabeza, reconocer unidades de acción facial y estimar la dirección de la mirada con código fuente disponible tanto para la ejecución como para el entrenamiento de los modelos.

Es una herramienta capaz de procesar las imágenes en tiempo real si se cuenta con el hardware adecuado y los resultados obtenidos en las primeras pruebas indican alto grado de precisión.

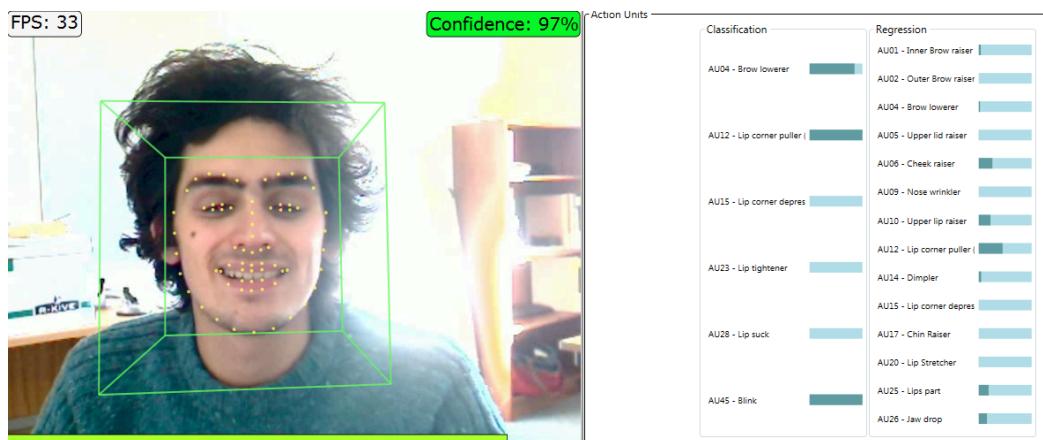


Figura 10: Interfaz gráfica de OpenFace mostrando Landmarks detectados junto con las unidades de acción y sus intensidades.

Fuente: OpenFace [<https://github.com/TadasBaltrusaitis/OpenFace/wiki>]

Como se aprecia en la figura 10, además de los landmarks OpenFace es capaz de calcular las unidades de acción. Al utilizar dos modelos para esto, uno brinda el nivel de activación de las mismas en una escala de 0 a 5, y otro simplemente 0 o 1 dependiendo si la unidad de acción se encuentra activa o no.

2.4.1.2 PyFeat

Pyfeat [52] es una herramienta avanzada diseñada para la detección y análisis de landmarks faciales. Este paquete de software de código abierto se destaca por su capacidad para identificar y extraer puntos clave del rostro humano con alta precisión y eficiencia.

Esta herramienta utiliza algoritmos avanzados para localizar puntos específicos en el rostro, como el contorno de la cara, las cejas, los ojos, la nariz y la boca. Es capaz de detectar y devolver 68 landmarks faciales. Estos puntos son esenciales para el análisis detallado de las expresiones faciales y para tareas posteriores de procesamiento de imágenes.

Puede ser utilizado inicializando el modelo detector de pyfeat que está preentrenado para identificar puntos clave en el rostro humano.

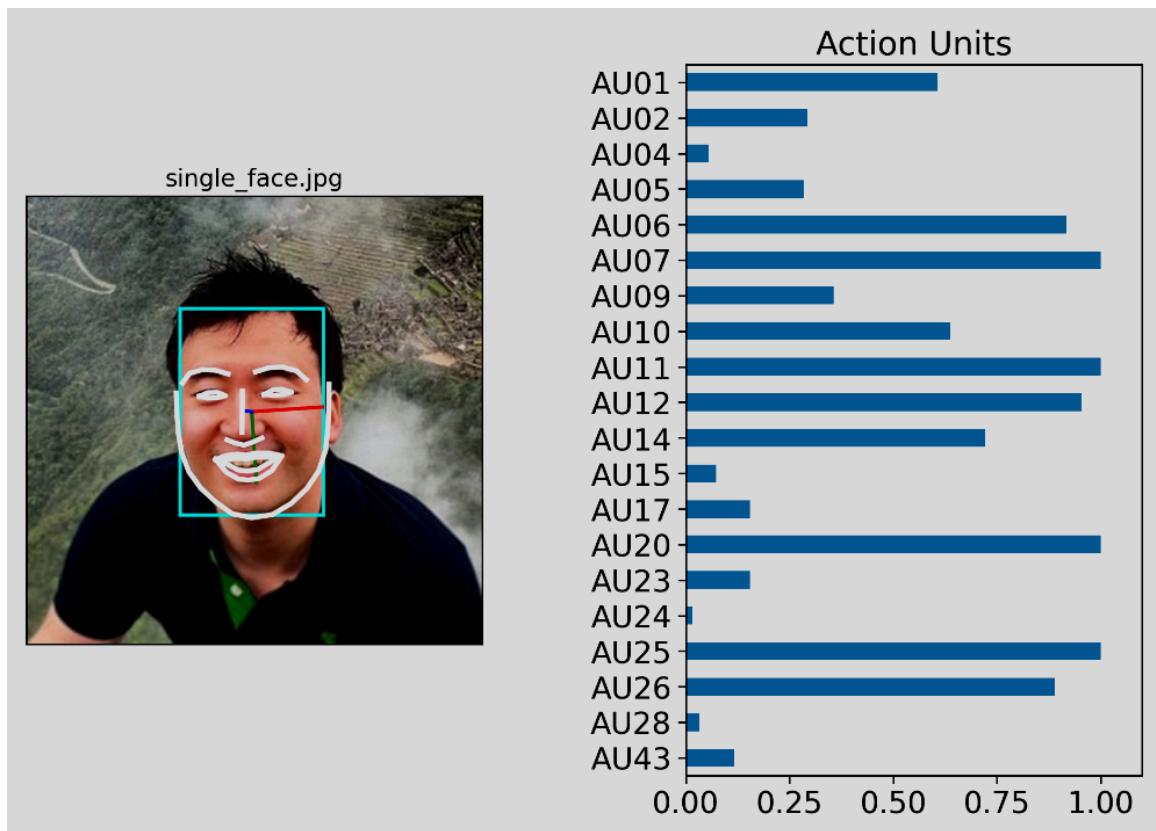


Figura 11: localización de landmarks y unidades de acción con la herramienta PyFeat

Fuente: Pyfeat [<https://py-feat.org/pages/intro.html>]

Como se puede observar en la figura 11, PyFeat no sólo calcula los landmarks, sino que también permite calcular las unidades de acción. En particular, proporciona 20 unidades de acción en una escala de 0 a 1, donde un valor de 0 indica que la unidad de acción no está presente y un valor de 1 indica que está completamente presente.

2.4.1.3 Face ++

Face++ es una empresa china especializada en tecnologías de reconocimiento facial y visión por computadora. La compañía se ha destacado en el desarrollo de algoritmos y sistemas para la detección y el análisis de rostros humanos en imágenes y videos.

La herramienta que brinda la posibilidad de reconocimiento facial se denomina “Detect API” que además de detectar rostros en una imagen, es capaz de reconocer componentes de los rostros como el contorno, ojos, nariz, etc.... [14]

Por otro lado, para tareas que requieran más precisión existe “Dense Facial Landmark API” que es capaz de obtener hasta 1000 landmarks de un rostro.

“Detect API” brinda la posibilidad de ser configuradas para localizar 83 o 106 landmarks faciales.

Puede ser consumido mediante una API o utilizando un SDK y provee una versión gratuita, con una cantidad limitada de peticiones por segundo, y una paga para requerimientos más demandantes.

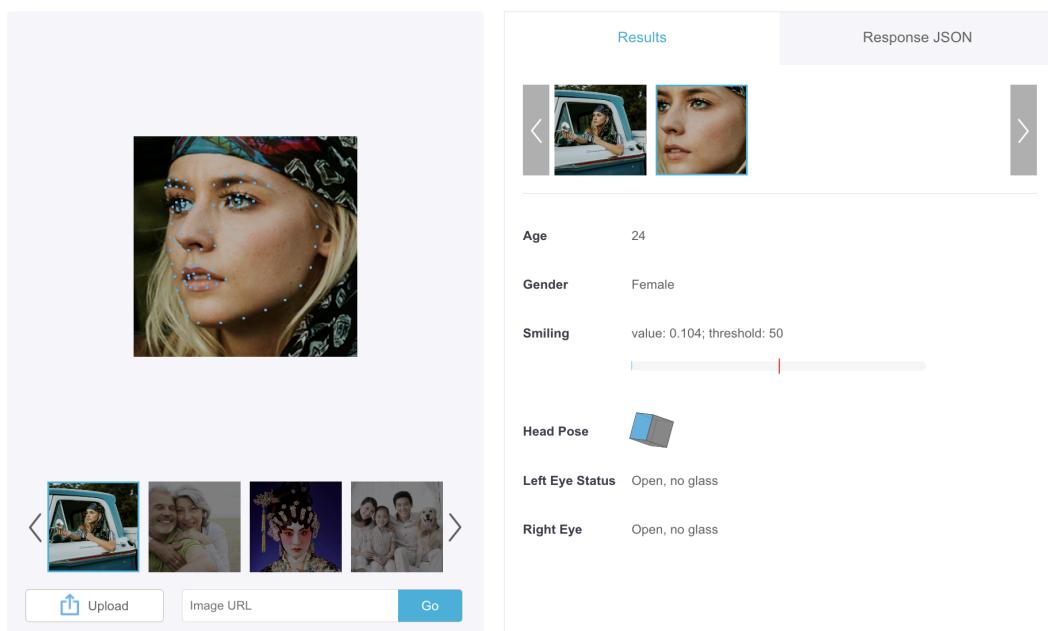


Figura 12: Localización de landmarks faciales con la herramienta "Detect API" de la plataforma Face++.

Fuente: Face++ Online [<https://www.faceplusplus.com/emotion-recognition/>]

2.4.2 Servicios para detección de emociones

A continuación se detallan servicios y herramientas para poder detectar emociones a partir de imágenes y/o videos de rostros humanos.

2.4.2.1 Face ++

Además de los servicios mencionados anteriormente, la plataforma Face++ provee un servicio denominado “Emotion Recognition” [14] (incluido en la herramienta “Detect API”) el cual es capaz de detectar 7 emociones: enojo, asco, miedo, felicidad, neutral, tristeza y sorpresa (anger, disgust, fear, happiness, neutral, sadness, surprise) dando un porcentaje de confianza de cada una.

Al igual que los demás servicios de la plataforma puede ser consumido mediante una API o utilizando un SDK y provee una versión gratuita y una paga.

2.4.2.2 ONNX FER+ Emotion Recognition

Open Neural Network Exchange (ONNX) es un formato de intercambio de modelos de aprendizaje automático de código abierto que permite a los desarrolladores representar modelos de aprendizaje automático de diferentes frameworks (como TensorFlow, PyTorch y Caffe) en un formato común y portátil.

FER+ Emotion Recognition (Facial Emotion Recognition) [15] es un modelo específico para la detección de emociones en imágenes faciales. La combinación de ONNX y FER+ se utiliza para representar y desplegar modelos de reconocimiento de emociones faciales en una variedad de plataformas y entornos. Este modelo pre entrenado es una red neuronal convolucional profunda entrenada sobre el dataset FER+ [16] y dada una imagen estima la probabilidad de cada una de las 7 emociones básicas más la emoción neutral.

ONNX [17] provee una librería que puede ser utilizada en diversos lenguajes, incluyendo python, y es de uso gratuito.

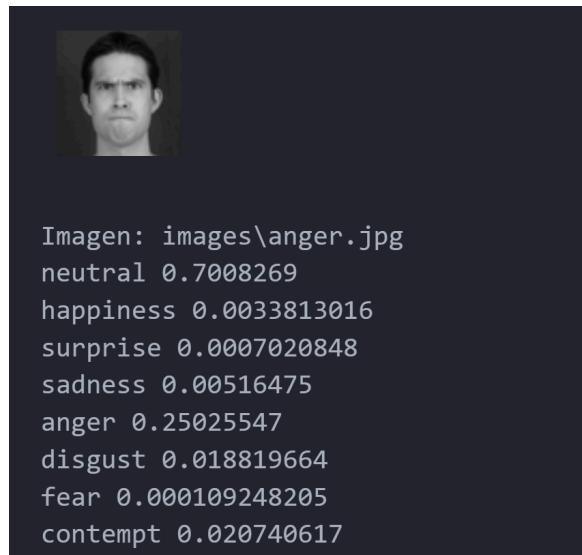


Figura 13: Ejemplo de corrida de una imagen usando ONNX FER+.

En la figura 13 se presenta un ejemplo de la ejecución del modelo FER+ aplicado a una imagen que representa la emoción de enojo. En su respuesta se puede observar las 7 emociones básicas acompañadas por el nivel de probabilidad otorgado por el modelo, principalmente detectar la emoción neutral y enojo.

2.4.2.3 PyFeat

PyFeat es un avanzado servicio de software que proporciona herramientas para la detección y análisis de emociones a partir de características faciales. Este servicio se basa en algoritmos de inteligencia artificial y aprendizaje automático para identificar y analizar expresiones faciales, lo que lo hace una opción ideal para nuestro sistema de detección de emociones en imágenes y videos.

Pyfeat utiliza modelos de aprendizaje automático para clasificar las emociones expresadas en la imagen o video. Estas emociones pueden ser clasificadas en categorías discretas (alegría, tristeza, miedo, ira, sorpresa, asco, neutra).

PyFeat provee una librería en python gratuita, open source y fácil de utilizar.

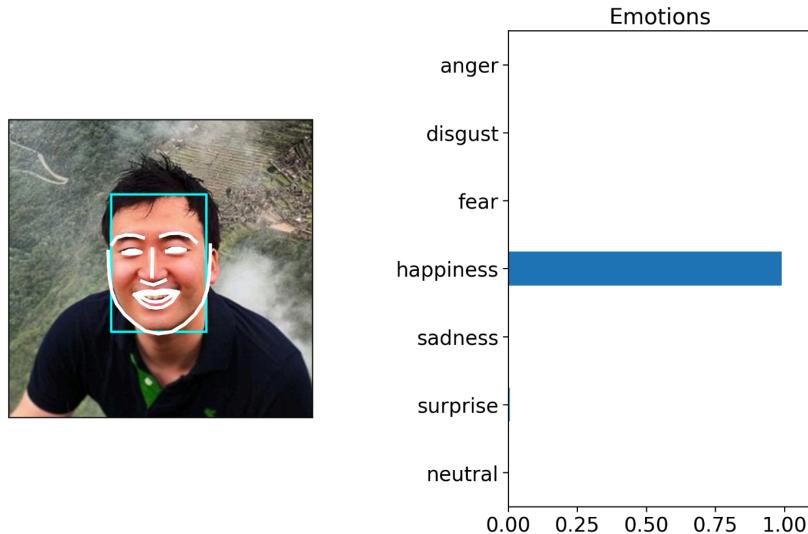


Figura 14: Ejemplo de predicción de Pyfeat para las emociones

Fuente: <https://py-feat.org/pages/intro.html>.

En la figura 14, se presenta un ejemplo de predicción de una imagen, en el cual se muestra la probabilidad asociada a cada emoción identificada en dicha imagen. En este caso particular, se trata de una persona sonriente, donde el modelo predice que la emoción predominante es la felicidad. La figura destaca la distribución de probabilidades para diversas emociones, confirmando que la mayor probabilidad corresponde a la emoción de felicidad.

2.4.2.4 AWS Rekognition

Amazon Rekognition [18] es uno de los servicios de AWS (Amazon Web Services) y ofrece diferentes herramientas para el análisis de imágenes y videos, entre ellos detección de emociones en imágenes de caras.

El servicio se puede consumir vía API o utilizando una librería, es de pago pero se ofrece una versión gratuita que dura 12 meses con una cantidad limitada de análisis mensuales.

La respuesta contiene las siete emociones básicas más la neutral y su nivel de confianza. Además brinda información adicional como los puntos que delimitan el rostro, landmarks con las coordenadas de los ojos, boca, nariz, etc., y otros atributos faciales como el género, anteojos, entre otros. [19]

3. Problema detectado y/o faltante

Como se mencionó anteriormente, hoy en día existen diversas herramientas que utilizan el modelo categórico que para una imagen dada tienen como salida una clasificación de las expresiones faciales propuestas por Paul Ekman [4] asociadas a un valor de intensidad entre 0 y 1 donde "0" significa que la expresión está ausente y "1" significa que está totalmente presente.

Este modelo, que clasifica las emociones en categorías discretas y específicas, tiene limitaciones que han generado críticas y desafíos en la comprensión completa del espectro emocional humano. Sus principales desventajas incluyen[29][30]:

1. La falta de variedades emocionales: al dividir las emociones en un conjunto discreto de categorías, el modelo categórico no logra capturar la riqueza y la complejidad de las emociones humanas. La vida emocional no se ajusta necesariamente a un conjunto limitado de etiquetas, lo que limita su capacidad para representar emociones mixtas, ambiguas o transicionales.

2. Limitación en la representación de la interrelación entre emociones: Este enfoque no considera la relación y la interacción entre diferentes estados emocionales. Las personas pueden experimentar emociones combinadas o influenciadas unas por otras, y el modelo categórico no refleja esta complejidad interconectada.

3. Dificultad para identificar emociones ambiguas: Las emociones a menudo no se ajustan a una sola categoría, lo que hace más difícil su clasificación dentro del modelo categórico. Las emociones ambiguas o complejas no se pueden capturar adecuadamente, lo que limita su aplicabilidad en situaciones donde las emociones no son claramente discernibles.

Es aquí donde el modelo dimensional puede permitirnos describir y comprender las emociones humanas en un espacio multidimensional, permitiendo una representación más completa y precisa de la diversidad emocional. Este se basa, como se mencionó anteriormente, en dos dimensiones principales: la valencia y el arousal. Para poder calcular estas dos dimensiones vamos a partir del modelo categórico utilizando puntualmente la intensidad de las emociones para obtener la valencia, y las unidades de acción calculadas a partir de los landmarks para obtener el arousal aplicando el algoritmo de Noldus. [22]

Hemos detectado que no existen herramientas accesibles que ofrezcan ambos modelos (categórico y dimensional) al momento de detectar emociones dentro de un contexto dado y de uso ilimitado (código abierto). Particularmente no existen aplicaciones visuales accesibles a todo público donde un individuo pueda utilizarse para obtener conclusiones de manera conjunta y sintetizada sobre las emociones por la que pasa una persona ante un contexto que

Herramienta para la evaluación de emociones en contextos abiertos

actúa como estímulo emocional, que puede tratarse de una videoconferencia, clase académica o cualquier contexto de reunión remota, que hoy en día en su mayoría no incluyen un sistema que les proporcione detectar las emociones.

4. Solución implementada

En esta sección detallaremos la solución implementada, haciendo un repaso por los componentes del sistema, arquitectura y detalles técnicos.

4.1 Introducción

La solución implementada, en líneas generales, consiste en un sistema distribuido capaz de detectar rostros y las emociones presentes en los mismos, ya sea a través de videos o imágenes mediante el uso de modelos de machine learning. Estas emociones se basan en el modelo categórico propuesto por Paul Ekman [4] que nos brinda una serie de 7 emociones junto a su valor de intensidad. A su vez, este modelo, luego sería utilizado por el modelo dimensional propuesto por James A. Russell [10]. Para éste último se utilizará la intensidad de las emociones proporcionada por el modelo anterior junto a las unidades de acciones calculadas por landmarks, para así poder aplicar el algoritmo de noldus [22] y poder determinar los valores de valencia y excitación. Como resultado, se generará un abanico de emociones disponibles para detectar en distintas situaciones.

Nuestro sistema se encuentra conformado por los siguientes componentes: procesadores de modelos de machine learning, una api y un joiner. Se decidió desacoplar el sistema en diferentes componentes comunicados entre sí debido a la necesidad de poder distribuir la alta carga de procesamiento que conlleva trabajar con imágenes y videos.

Así es como la solución busca abordar el problema utilizando lo que llamamos **procesadores**. Estos procesadores son los componentes que corren los modelos de Machine Learning (valence-processor y arousal-processor) capaces de reconocer rostros y características faciales presentes en los mismos. Nuestros procesadores son *stateless*, es decir, no guardan un estado, y esto nos brinda una gran ventaja, la escalabilidad. Dado que todas las comunicaciones entre componentes son asincrónicas, y se realizan utilizando colas de RabbitMQ[44], nuestro sistema tiene la capacidad de escalar horizontalmente, sumando nuevas réplicas de los procesadores, y de esta forma enfrentar un aumento en la carga de ser necesario.

Sin embargo, es importante mencionar que la arquitectura no es elástica, es decir, no escala automáticamente frente a un aumento en las solicitudes de los usuarios. El escalamiento debe realizarse de forma manual.

Por otro lado, como se mencionó, contamos con dos componentes adicionales: una *API* para recibir las solicitudes de los usuarios y enviar los datos a los procesadores, y un *Joiner* que se encarga de hacer un join entre los resultados de ambos procesadores para luego publicar los mismos.

Otro de los principales motivos por los cuales decidimos utilizar una comunicación asincrónica entre los diferentes componentes es el hecho de poder realizar un procesamiento

en *batches* (lotes de respuestas). De esta forma, el sistema es capaz de devolver resultados a medida que estos batches son procesados, y no es necesario esperar que todo el video se procese completamente para poder obtenerlos. Esto es muy conveniente para videos de larga duración, considerando que su procesamiento puede demorar un tiempo considerable y así el usuario puede tener una experiencia lo más cercana al tiempo real de el video a procesar.

A continuación un diagrama de arquitectura simplificado:

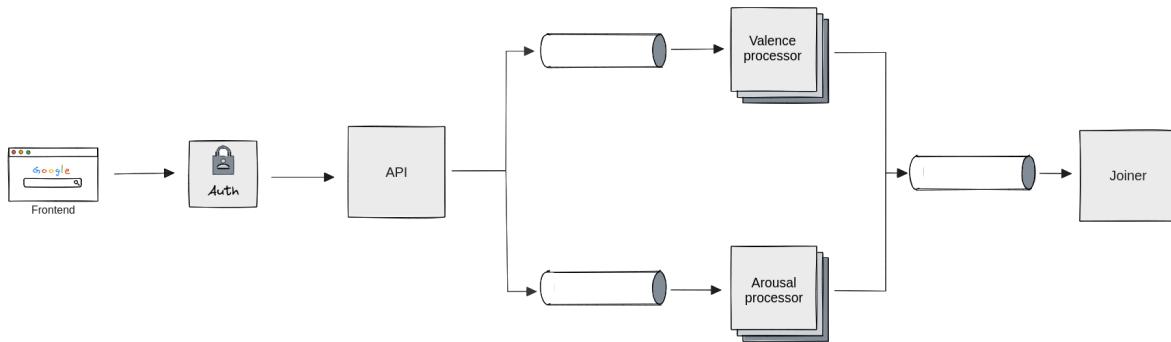


Figura 15: Diagrama de arquitectura simplificado.

4.2 Detalle de los componentes

A continuación, se enumeran y describen en detalle los componentes individuales que conforman a nuestro sistema

4.2.1 API

El componente *API* es el punto de entrada a nuestro sistema. El mismo provee una API Rest [49] con un conjunto de endpoints que serán utilizados por la interfaz web.

A gran escala, la API se divide en dos grandes bloques. En primer lugar, disponemos de un bloque que se encarga del manejo de usuarios y la autenticación de los mismos. Luego, tenemos el bloque encargado del procesamiento de los inputs multimedia, la lectura posterior de su análisis y la lectura de aquellos ítems ya procesados y almacenados.

Este componente fue implementado en Python utilizando la librería FastAPI [45] por dos motivos principales: es un framework sencillo de utilizar, que además brinda una documentación OpenAPI (*/docs*) interactiva lo cual facilita el proceso de desarrollo, y por otro lado, necesitábamos un framework capaz de realizar operaciones asíncronas, principalmente para poder realizar tareas en segundo plano, logrando así no ‘bloquear’ a los usuarios en peticiones que de forma sincrónica podrían demorar mucho tiempo, por ejemplo, subir y procesar un video.

Herramienta para la evaluación de emociones en contextos abiertos

A continuación un diagrama de secuencia del proceso (simplificado) de subir un video a la API:

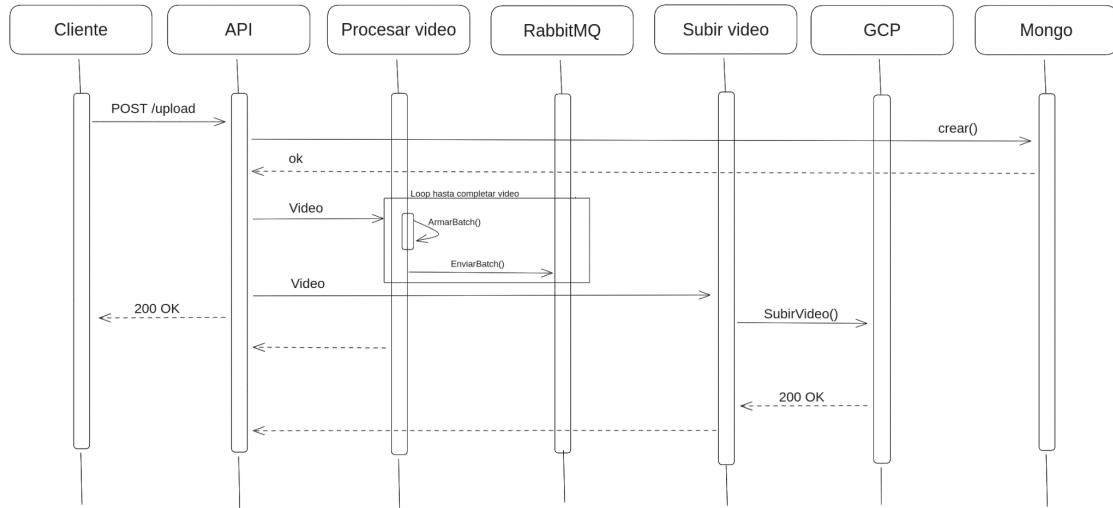


Figura 16: Diagrama de secuencia: subir vídeo.

Como se ve en el diagrama de la figura 16, una vez que el cliente sube un video se lanzan tareas en ‘background’ (segundo plano) las cuales van a procesar el video. Una de las tareas arma los lotes de frames y los envía a RabbitMQ, mientras que la otra sube el video a Google Cloud Platform Object Storage.

De esta manera se puede ver cómo devolvemos una respuesta al cliente (para no bloquearlo y que deba esperar a que se complete el procesamiento) de forma que pueda comenzar a solicitar batches procesados y, en el mientras tanto, continuamos procesando el video subido.

Cabe destacar que para el caso de imágenes este proceso es síncrono, es decir, el usuario subirá una imagen y nuestra respuesta ya contendrá toda la información de la imagen procesada.

Para lograr este procesamiento asincrónico utilizamos como herramienta principal la librería `asyncio` [35], la cual facilita la concurrencia y paralelismo al permitir que múltiples tareas se ejecuten de forma cooperativa en un solo hilo de ejecución.

Por otro lado nos vimos en la necesidad de buscar librerías que soporten la ejecución asíncrona, entre ellas `aiormq` [36] para RabbitMQ, `Motor` [37] para MongoDB y `cloud-aio` [38] para interactuar de forma asíncrona con el servicio Object Storage de Google Cloud Platform.

Además de lo mencionado anteriormente, la API expone otros endpoints que permiten operaciones como buscar información de un batch por ID, buscar la información de un batch por tiempo, cargar estímulos para determinados videos o imágenes, entre otros.

4.2.2 Procesadores

Como se mencionó anteriormente, el sistema cuenta con dos componentes procesadores los cuales funcionan de forma similar como se puede observar en la figura 17.

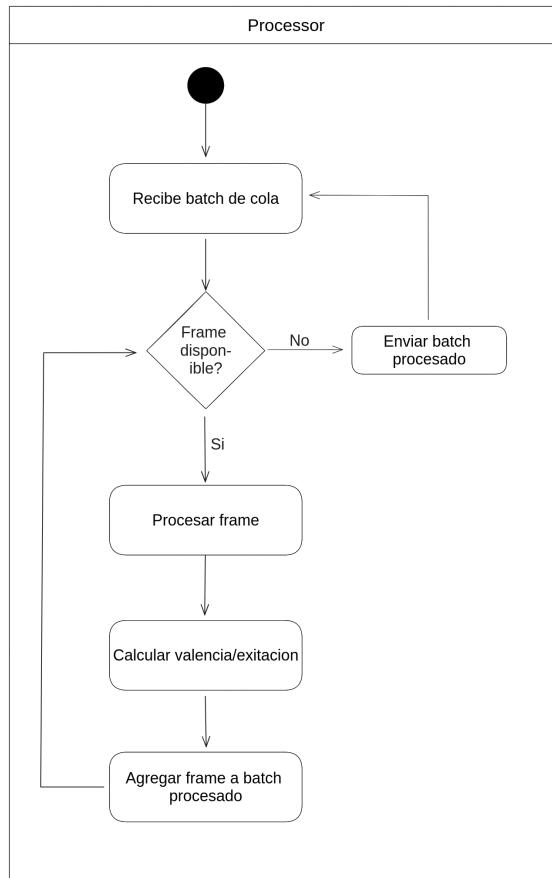


Figura 17: Diagrama de actividad de los procesadores.

Los pasos en donde difieren los procesadores son '*procesar frame*', ya que utilizan diferentes modelos, y en '*Calculo valencia/excitación*', ya que el '*'arousal-processor'*' se encarga de la excitación y el '*'valence-processor'*' de la valencia.

A continuación una explicación en detalle de cada procesador.

4.2.2.1 Valence processor

Este componente es responsable de calcular la valencia a partir de una imagen. El cálculo de la valencia se basa en el modelo de Noldus [22] mencionado anteriormente, utilizando las siete emociones básicas planteadas por el psicólogo Paul Ekman. La valencia se calcula como

la intensidad de la emoción 'feliz' menos la intensidad de la expresión negativa con mayor intensidad.

Actualmente, el sistema posee dos modelos para calcular la valencia: FerPlus y PyFeat.

Por un lado, **FerPlus** es un modelo de aprendizaje profundo diseñado para la detección y clasificación de emociones a partir de imágenes faciales. Este modelo se basa en una arquitectura de red neuronal convolucional (CNN) y ha sido entrenado con el conjunto de datos FER+ (Facial Expression Recognition Plus), una extensión del conjunto de datos FER (Facial Expression Recognition) con anotaciones más precisas. El modelo clasifica las imágenes en una de las ocho categorías emocionales: Neutral, Feliz, Triste, Sorprendido, Asustado, Disgustado, Enojado y Desprecio. Procesa una imagen de entrada de un rostro y produce un vector de probabilidades para cada una de las emociones mencionadas, considerando la emoción con la probabilidad más alta como la predominante en la imagen.

Por otro lado, el modelo **PyFeat** es una herramienta avanzada para el análisis de expresiones faciales, diseñada para la detección de emociones básicas según las teorías de Paul Ekman. Utiliza redes neuronales convolucionales (CNN) y algoritmos de visión por computadora para identificar y clasificar seis emociones universales: felicidad, tristeza, sorpresa, miedo, ira y disgusto. Además de las emociones, PyFeat puede detectar unidades de acción (AUs) del sistema de codificación de acción facial (FACS), proporcionando un análisis detallado de los movimientos musculares faciales.

Aunque el sistema está adaptado para utilizar ambos modelos, se ha decidido realizar todas las pruebas utilizando el modelo PyFeat. A pesar de que este modelo puede tardar más en proporcionar una respuesta en comparación con el modelo FerPlus, ofrece una mayor precisión en la determinación de la emoción predominante/

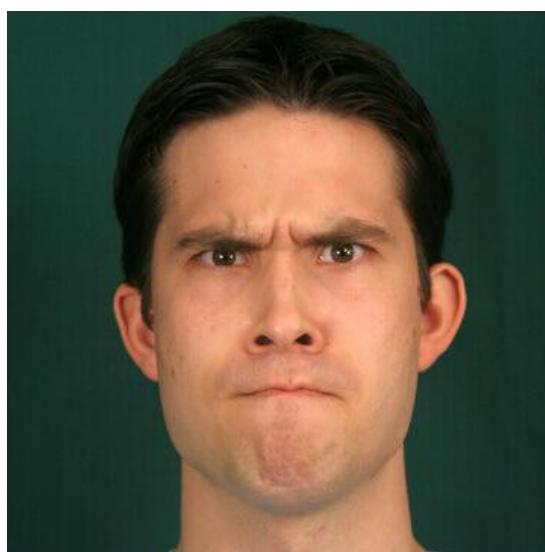


Figura 18: Imagen de actor presentando la emoción enojo.

Modelo	FerPlus	PyFeat
Enojo	0,1905854000	0,9775949121
Tristeza	0,0125319510	0,0002306037
Asco	0,0000711234	0,0004401084
Miedo	0,0020219786	0,0001756316
Felicidad	0,0060854750	0,0059974161
Sorpresa	0,0004126069	0,0001034793
Neutral	0,7668633000	0,0154577401

Tabla 1: Comparativa de resultados modelo Ferplus vs PyFeat

La tabla 1 muestra la comparativa entre los dos modelos explorados para el cálculo de emociones básicas del modelo categórico para el caso de la figura 18 donde muestra el rostro de un actor que corresponde a la emoción de enojo. Se puede observar que para el modelo de pyfeat se obtienen mejores resultados que para el modelo de ferplus que en general le da una mayor probabilidad a la emoción neutral. Para una revisión más exhaustiva y detallada de todas las comparativas, puede acceder al spreadsheet con el resto de los datos y análisis mediante el siguiente enlace: [Acceso al Spreadsheet](#). Este documento contiene comparativas adicionales que permiten una mejor comprensión de las diferencias de rendimiento entre ambos modelos en la detección de diversas emociones.

El flujo de este componente consiste en recibir un lote de un video determinado a través de una cola denominada “valence_frames” de RabbitMQ, calcula la valencia para todas las imágenes dentro del lote y luego envía estos resultados a otra cola denominada “processed”.

4.2.2.2 Arousal processor

Este componente se encarga de calcular el valor de excitación utilizando las Unidades de Acción.

Para poder obtener las unidades de acción a partir de las imágenes de rostros utilizamos el modelo OpenFace [39], el cual nos devuelve la intensidad de activación de un conjunto de unidades de acción. Particularmente OpenFace es capaz de reconocer las siguientes unidades

de acción: AU1, AU2, AU4, AU5, AU6, AU7, AU9, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU28 y AU45.

Para poder integrar OpenFace a nuestro sistema, fue necesario realizar modificaciones sobre el código existente. La necesidad más grande era poder recibir imágenes por una cola de RabbitMQ en lugar de que el usuario cargue las mismas directamente en OpenFace (como originalmente funciona), es por eso que se implementó un nuevo componente llamado '*RabbitCapture*', el cual reemplaza al componente original '*ImageCapture*', y de esta forma pudimos suplir esa necesidad.

Como se detalló en el estado del arte, hacemos uso de esta intensidad que oscila entre 0 y 5 (siendo “0” inactiva y “5” completamente activada). Para determinar la excitación utilizamos las 5 unidades de acción con mayor valor de activación dentro del conjunto y calculamos el promedio. Este promedio luego se normaliza para llevarlo a una escala de [-1, 1]. En cuanto a las unidades de acción, antes de enviar los resultados al Joiner, también se normalizan llevándolas a una escala de [0, 1].

Previo a integrar OpenFace a nuestro sistema, se desarrolló el componente utilizando la librería de FaceTorch, la cual internamente utiliza el modelo de OpenGraphAU[46] para el reconocimiento de unidades de acción, pero luego de realizar pruebas, comparando ambos modelos con Noldus [22], la definición de las imágenes y nuestra percepción personal, concluimos que ambos modelos eran muy precisos y brindaban resultados acertados.

Las pruebas que realizamos consistieron en tomar una imagen de la cual ya sabíamos la emoción predominante, por ejemplo la siguiente figura con la emoción ‘sorpresa’ (*surprise*), y comparar las unidades de acción con mayor valor de activación devuelto por ambos modelos. También comparamos los resultados de los modelos contra Noldus FaceReader [22], contra las unidades de acción que se deberían activar “por definición”[51] y a su vez contra las unidades de acción que nosotros creímos que deben activarse para dicha imagen (basándose en la definición de cada unidad de acción).



Figura 19: Imagen de actor presentando la emoción sorpresa.

Para la imagen de la figura 19 se obtuvo la siguiente tabla:

UA	OpenFace	FaceTorch	Noldus	Definición	Nosotros
1	X			X	X
2	X		X	X	X
3					
4			X		
5	X	X	X	X	X
6					
7		X			
9					
10					X
12					
14					
15					
17		X			
20					

UA	OpenFace	FaceTorch	Noldus	Definición	Nosotros
23					
25	X	X	X		X
26	X	X	X	X	X
28					

Tabla 2: Comparación de unidades de acción calculadas por los diferentes modelos.

Como se puede observar en la tabla 2, marcamos las 5 unidades de acción con mayor activación devueltas en cada modelo, las cuales son utilizadas para calcular la excitación como se explicó anteriormente. Se puede apreciar que los resultados son muy similares para todos los casos y es por eso que lo que finalmente nos llevó a elegir OpenFace por sobre FaceTorch (OpenGraphAU) no fue la precisión del modelo, sino que fue el tiempo de procesamiento, el cual era notoriamente inferior en el modelo elegido.

Las pruebas y valores para otras emociones pueden ser vistas en la siguiente tabla: [Tablas Comparativas Unidades De Acción](#).

A diferencia del algoritmo de Noldus [22], optamos por no calcular el promedio de los últimos 60 segundos ya que encontramos que no era necesario hacer una corrección con respecto al valor obtenido. De hecho, el arousal tenía a cero si el promedio de los últimos 60 segundos se mantenía muy cercano al valor de la excitación.

A continuación, en la figura 20, se puede observar un diagrama del cálculo del arousal simplificado.

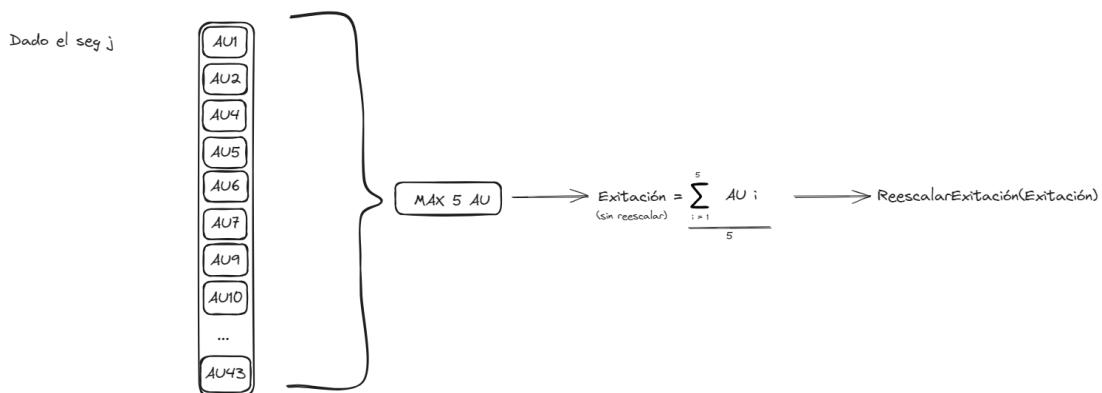


Figura 20: Ejemplo de cálculo del arousal en nuestro sistema.

4.2.3 Joiner

El componente *Joiner* es el último componente del pipeline de procesamiento. En líneas generales, este se encarga de recibir los batches procesados desde los procesadores (*arousal-processor* y *valence-processor*) a través de una cola de RabbitMQ, unirlos en un mismo objeto y enviarlo para poder ser consumido por la API.

Los batches recibidos contienen el ID del usuario, el número del batch, el ID del archivo al que pertenece el batch, y además el origen, siendo este último ‘*arousal*’ o ‘*valence*’. De esta forma, el Joiner cuenta con toda la información necesaria para fusionar estos batches.

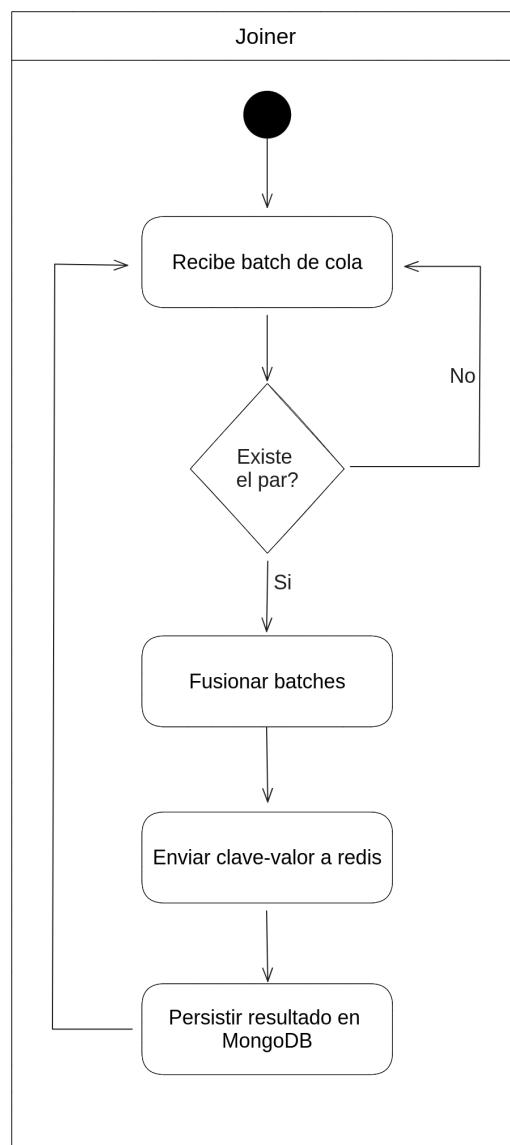


Figura 21: Diagrama de actividad del funcionamiento del joiner.

Como se ve en el diagrama anterior, el Joiner envía los batches fusionados a Redis para que sean consumidos por la API y también a MongoDB para persistirlos. Ambos casos se detallarán en la próxima sección.

4.2.4 Aplicación Web

Este componente representa la interfaz de usuario diseñada para facilitar la interacción del mismo con el sistema de detección de emociones en imágenes y videos. La aplicación se centra en la experiencia visual y la usabilidad del sistema, permitiendo al usuario cargar videos, imágenes, el estímulo asociado a ellos y visualizar los resultados del análisis emocional de manera clara y comprensible tanto para el modelo categórico como dimensional.

Dentro de esta sección se detallaremos una serie de funcionalidades principales que conforman a nuestro sistema. Entre ellas se encuentra la carga de videos e imágenes tanto del video a procesar como del estímulo asociado a dicho video, la sección de resultados del modelo categórico y dimensional, la sección de unidades de acción, etc.

4.2.4.1 Login

La interfaz de usuario tiene como objetivo permitirle al usuario acceder al sistema e incluye un formulario con dos campos principales: uno para el email y otro para la contraseña, donde el usuario puede acceder al sistema mediante la verificación de sus credenciales. En la figura 22 podemos observar la implementación de nuestro login. Sin credenciales no se puede acceder al sistema.

Detección de emociones

Login



Email *

Contraseña *

LOGIN

¿No tienes una cuenta? [Regístrate](#)

Figura 22: login del sistema de detección de emociones.

4.2.4.2 Registro

En caso de no contar con credenciales, esta interfaz facilita un formulario que permite a los nuevos usuarios crear una cuenta en el sistema mediante la introducción de un email y una contraseña. A su vez contiene el campo de confirmación de la contraseña.

Crea tu cuenta

Email *

Contraseña *

Confirmar contraseña *

REGISTRARSE

¿Ya tienes una cuenta? [Login](#)

Figura 23: registro del sistema de detección de emociones.

4.2.4.3 Nuevo Análisis Video

Esta pantalla permite realizar un nuevo análisis de un video seleccionado por el usuario. La interfaz se divide en dos secciones principales: la sección de carga de videos y la sección de resultados. La sección de carga de videos permite al usuario cargar el video deseado junto con su estímulo asociado. La sección de resultados incluye las emociones detectadas instantáneamente por el modelo de Ekman y el modelo de Russell, las unidades de acción identificadas, la valencia y la excitación a lo largo del tiempo, así como una sección de resumen. Cada una de estas secciones se detalla más adelante. En la figura 24 se presenta un ejemplo de esta pantalla.

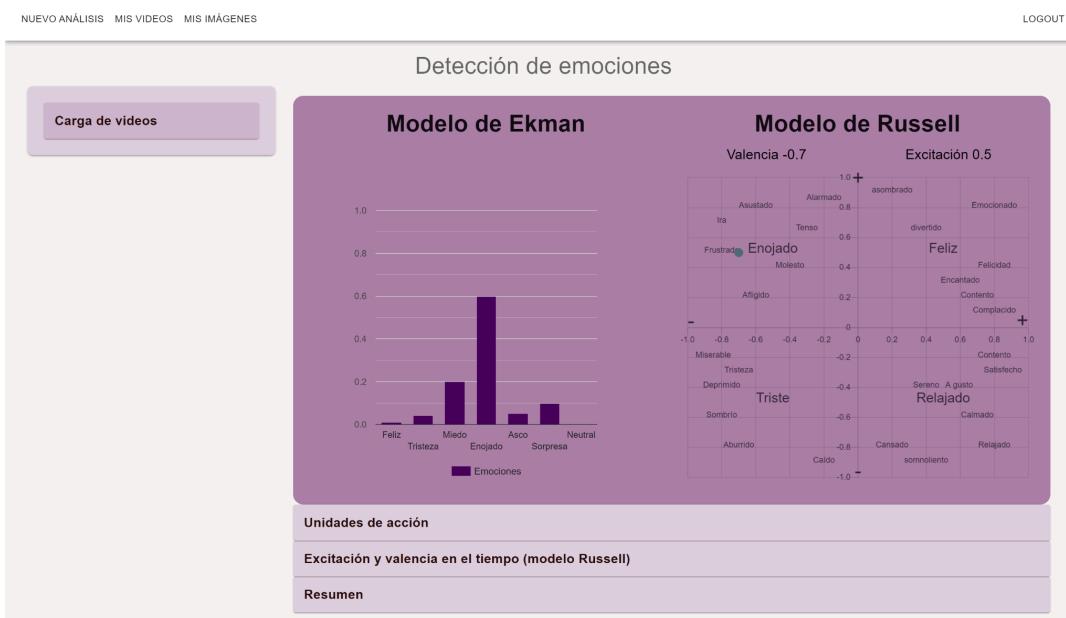


Figura 24: Página de inicio para un nuevo análisis de un video.

4.2.4.4 Nuevo Análisis Imagen

De manera similar a la pantalla anterior, esta interfaz se encarga de realizar un nuevo análisis, pero para imágenes. La interfaz incluye una sección de carga de la imagen a procesar junto con su estímulo asociado. Además, presenta una sección de resultados que abarca las emociones detectadas en el rostro de la persona mediante el modelo de Ekman y el modelo de Russell, las unidades de acción identificadas, y una comparativa entre la valencia y la excitación esperadas y las calculadas. En la figura 25 se presenta un ejemplo de esta pantalla.

Herramienta para la evaluación de emociones en contextos abiertos

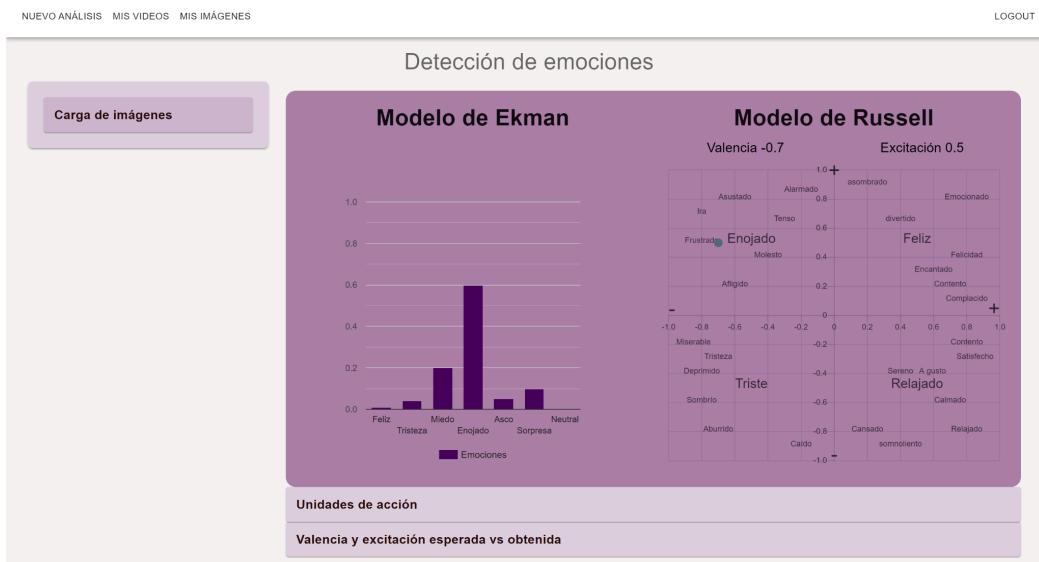


Figura 25: Página de inicio para un nuevo análisis de una imagen.

4.2.4.5 Mis Videos

La sección "Mis Videos" está destinada a mostrar los videos que el usuario ha procesado previamente. Si el usuario decide realizar un análisis de un video que ya ha sido procesado, puede acceder a esta sección y seleccionar el video deseado. De esta manera, será redirigido a la pantalla de análisis, donde podrá visualizar el video y los resultados correspondientes. La figura 26 ilustra un ejemplo de esta funcionalidad.

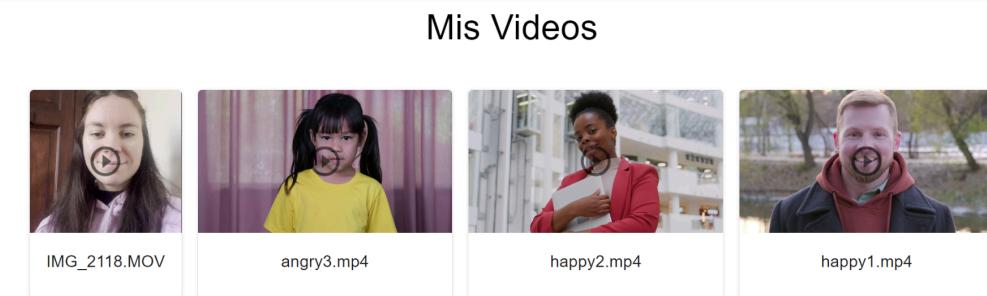


Figura 26: Sección mis videos ya procesados por el sistema.

4.2.4.6 Mis Imágenes

La sección "Mis Imágenes" está destinada a mostrar las imágenes que el usuario ha procesado previamente. Si el usuario decide realizar un análisis de una imagen que ya ha sido analizada, puede acceder a esta sección y seleccionar la imagen deseada. De esta manera, será

Herramienta para la evaluación de emociones en contextos abiertos

redirigido a la pantalla de análisis, donde podrá visualizar la imagen y los resultados correspondientes. La figura 27 ilustra un ejemplo de esta funcionalidad.



Figura 27: Sección mis imágenes ya procesadas por el sistema.

4.2.4.7 Sección Modelo Ekman y Modelo Russel

Esta sección se encarga de mostrar las emociones detectadas para un instante dado del video que se está procesando. A medida que el video se reproduce, la información se actualiza en una unidad de tiempo de un segundo. En la figura 28, se observa que, por un lado, se presenta el gráfico del modelo de Ekman, el cual muestra la probabilidad de ocurrencia de las siete emociones básicas (felicidad, tristeza, enojo, asco, sorpresa y neutral) en el rostro de la persona. Por otro lado, se incluye el gráfico del modelo de Russell, que representa la valencia y la excitación instantáneas en un eje de coordenadas. Este gráfico también indica en qué cuadrante emocional se clasifica a la persona (enojado, feliz, triste, relajado) y, con mayor detalle, a qué emoción específica corresponde. En total, el modelo de Russell puede determinar hasta 28 emociones diferentes.

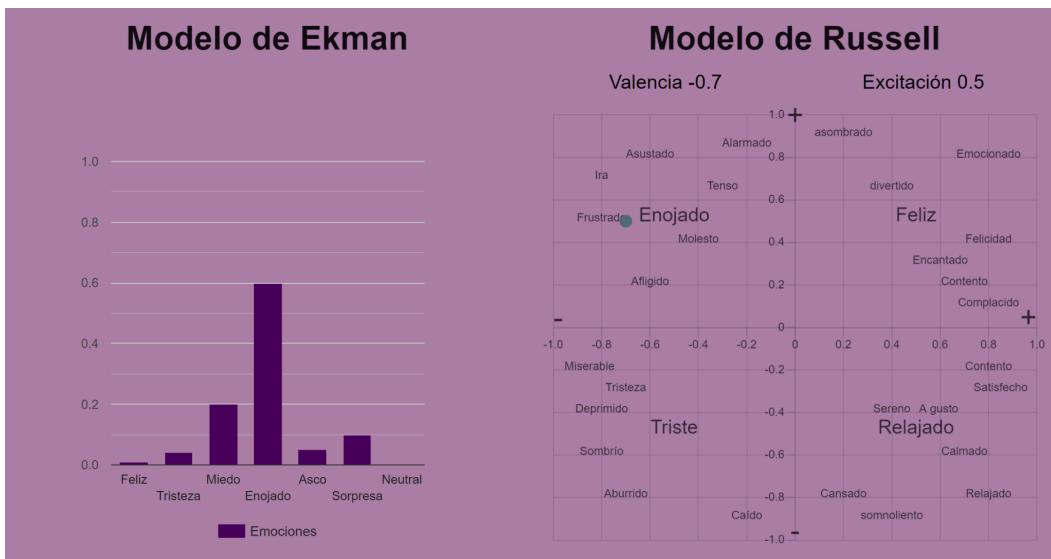


Figura 28: Detección de emociones modelo de Ekman y de Russel.

4.2.4.8 Sección carga de video a procesar y video estímulo

En esta sección, se permite cargar el video de la reacción de una persona que se desea procesar, así como el estímulo asociado, si lo hubiera. Además, el usuario puede cargar la valencia y la excitación promedio del video de estímulo, lo cual será utilizado posteriormente para comparar los promedios esperados con los obtenidos por el sistema. Es importante destacar que esta valencia y excitación promedio deben estar en una escala de -1 a 1.

Al momento de subir el video, tanto el video a procesar como el video de estímulo se sincronizan entre sí, lo que permite un análisis más detallado. Esto se debe a que, si se detiene el video, es posible observar las emociones detectadas, la valencia y la excitación estimadas para ese instante de tiempo, así como el estímulo al que la persona estaba siendo expuesta en ese momento. La figura 29 muestra un ejemplo de la carga de un video y un estímulo, junto con su valencia y excitación promedio.

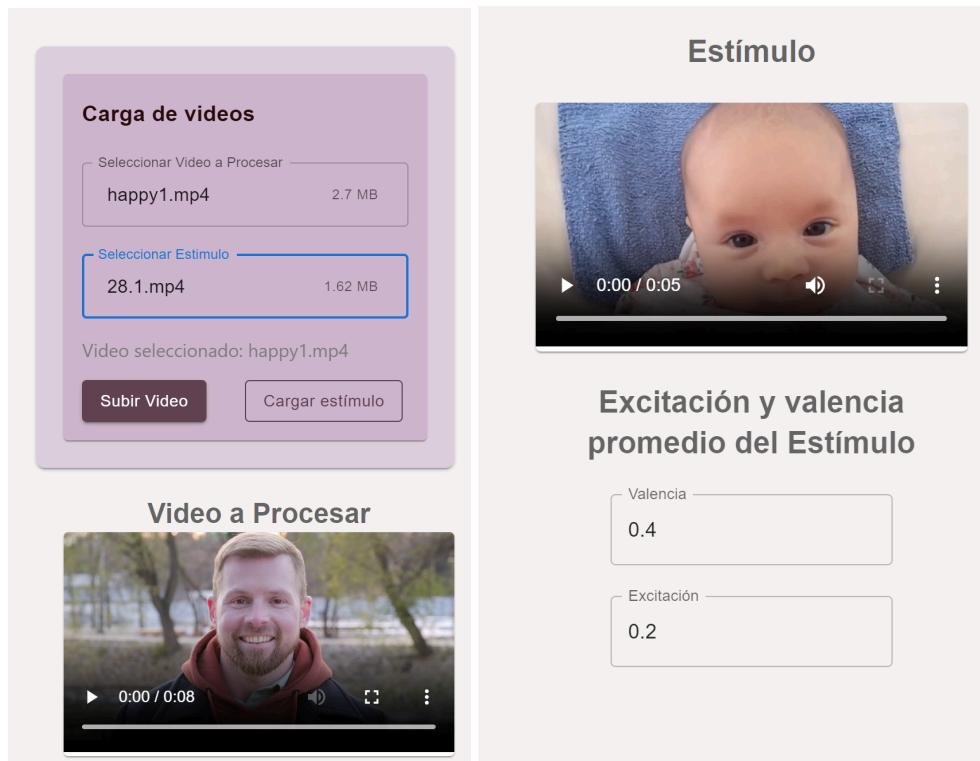


Figura 29: Ejemplo de carga de video a procesar junto a su estímulo.

4.2.4.9 Sección Carga de imagen a procesar y estímulo

En esta sección, se permite cargar la imagen de la reacción de una persona que se desea procesar, así como el estímulo asociado, si lo hubiera. Además, el usuario puede cargar la valencia y la excitación de dicho estímulo lo cual será utilizado posteriormente para comparar los valores esperados con los obtenidos por el sistema. Es importante destacar que esta valencia y excitación deben estar en una escala de -1 a 1.

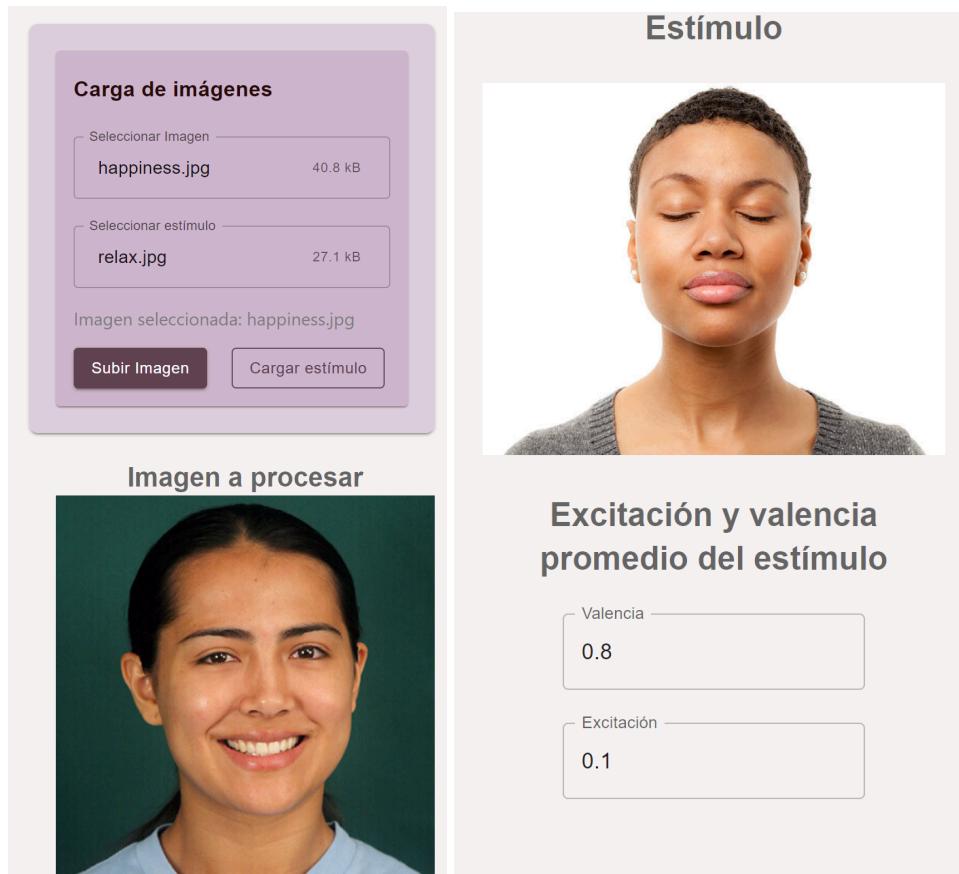


Figura 30: Ejemplo de carga de imagen a procesar junto a su estímulo.

4.2.4.10 Sección unidades de acción

El objetivo de esta interfaz es mostrar el nivel de activación de las unidades de acción que presenta la persona en cada momento. En total, se muestran 17 unidades de acción, las cuales están relacionadas con la parte superior e inferior del rostro.

Las unidades de acción se miden en una escala de 0 a 1, donde 0 indica que la unidad de acción no está presente en el rostro y 1 indica que está completamente presente. En la figura 31 se presenta un ejemplo de la sección de unidades de acción para un escenario específico de una persona.

Herramienta para la evaluación de emociones en contextos abiertos

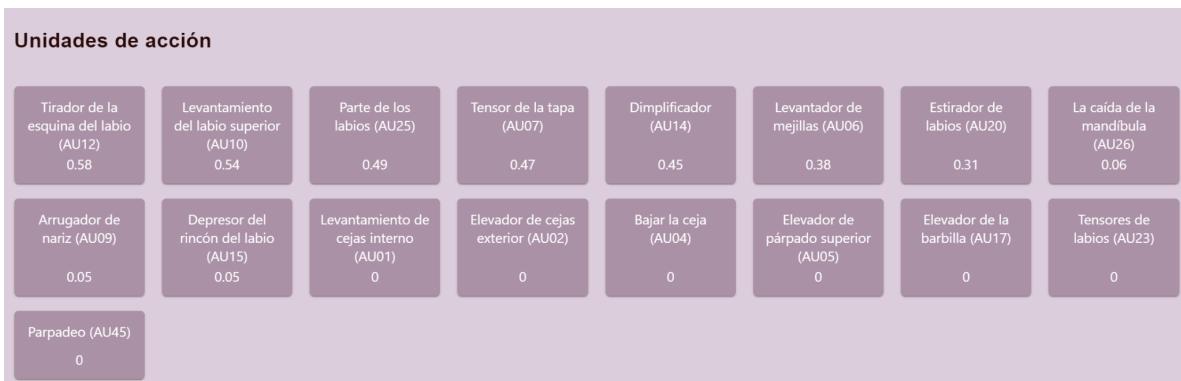


Figura 31: Ejemplo de la sección de unidades de acción.

4.2.4.11 Sección Excitación y Valencia en el Tiempo

El objetivo de esta sección es visualizar la valencia y la excitación en función del tiempo, lo que permitirá realizar un análisis más integral al observar la variación de los valores de excitación y valencia por los que la persona del video ha pasado. Este gráfico se actualiza continuamente a medida que el video se procesa.



Figura 32: Sección de excitación y valencia en el tiempo.

En la figura 32 se muestra un ejemplo de la interfaz de la sección que presenta la variación de excitación y valencia en el tiempo para un video dado.

4.2.4.12 Sección Resumen

Esta sección ofrece un resumen integral del análisis realizado, proporcionando una visión detallada de la excitación y valencia promedio de la persona, así como una representación visual mediante una nube de puntos en el circunflejo del modelo de Russell. Además, en caso de que el usuario cargue la valencia y excitación promedio del estímulo esperado, se permite una comparación para verificar si los valores calculados por el sistema coinciden con el cuadrante correspondiente en la nube de puntos.

Este análisis es de gran utilidad, ya que facilita la visualización del rango de emociones a lo largo del video de manera integrada. En la figura 33 se presenta un ejemplo donde la valencia y excitación promedio del estímulo corresponde al cuadrante "relajado". La nube de puntos se encuentra principalmente dispersa en dicho cuadrante, y la excitación y valencia promedio calculadas por el sistema están cercanas en valores a las esperadas por el estímulo.

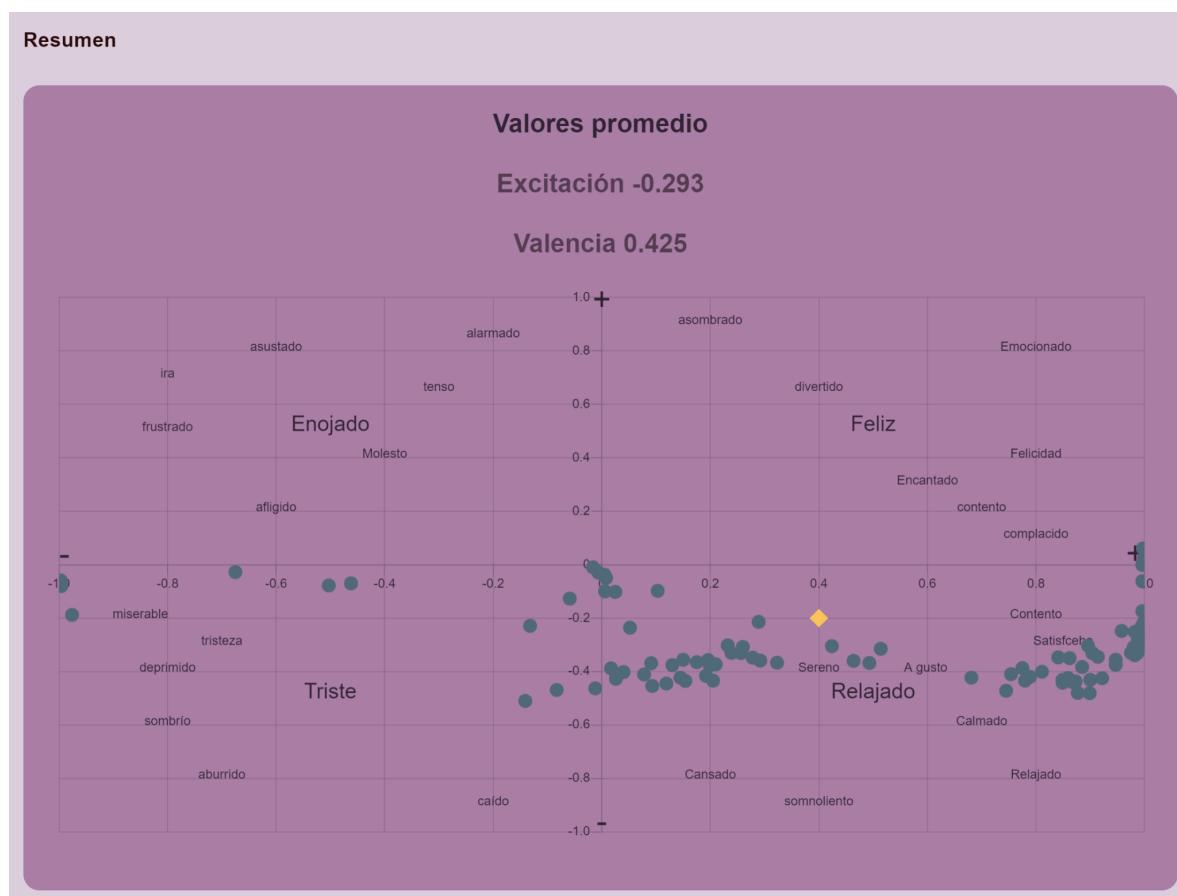


Figura 33: Sección de excitación y valencia en el tiempo.

4.3 Arquitectura y Especificaciones técnicas

En esta sección presentamos detalladamente decisiones y aspectos que planteamos en nuestro sistema al definir la arquitectura.

4.3.1 Registro y login de usuarios

El sistema debía ser capaz de procesar datos de diferentes clientes en simultáneo, y como mencionamos anteriormente, nuestros procesadores no guardan un estado, es por eso que nos vimos en la necesidad de agregar un identificador único en el sistema para cada usuario, de forma que cada batch de información que es procesado cuenta con este identificador.

Además, este identificador será utilizado para lograr la comunicación entre el Joiner y la API, como también para poder persistir los videos e imágenes procesados para un futuro análisis de los resultados. En las siguientes secciones detallaremos ambos casos.

Dicho esto, decidimos implementar un sistema de Registro e Inicio de sesión utilizando *JWT (JSON Web Tokens)* [40], de esta manera solo usuarios autenticados pueden utilizar el sistema y además cada usuario solo podrá ver sus videos e imágenes procesados, no los de otros usuarios del sistema.

El registro e inicio de sesión se realiza utilizando contraseña y email, siendo este último el que usaremos como `user_id` para todas las operaciones de dicho usuario.

4.3.2 Comunicación Joiner-API

Como mencionamos en la introducción y en el detalle de los componentes, nuestro sistema es capaz de publicar los resultados a medida que estos están disponibles, sin necesidad de esperar a que todo el video se procese.

Para lograr esta comunicación entre los resultados del joiner y la API, la cual necesitamos que sea lo más rápido posible, decidimos utilizar Redis. De esta forma, aseguramos buena performance (dado que redis está diseñado para esto) y también nos abstraemos de los inconvenientes de concurrencia que puede implicar esta comunicación.

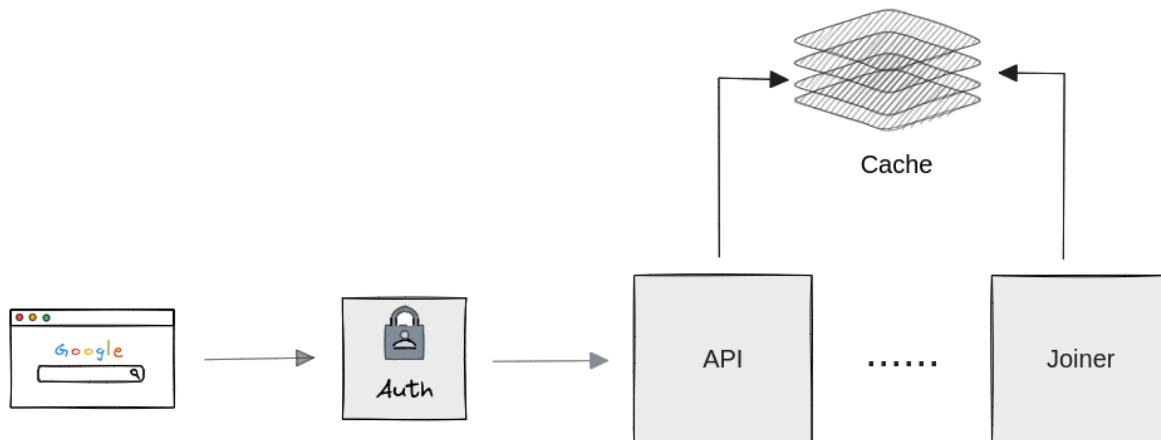


Figura 34: forma de comunicar resultados del Joiner a la API.

Es por eso que el joiner, a medida que recibe los batches procesados desde los procesadores, publicará en redis pares de *clave-valor*, donde estas claves son únicas por usuario y video en el sistema (compuestas por user_id, video_id, batch) y los valores son, justamente, la información de los batches procesados.

En la figura 34 podemos observar de manera ilustrativa como tanto la API como el Joiner apuntan al mismo Caché, que en este caso sería Redis.

De esta forma, cuando los clientes soliciten datos de ciertos batches a la API, esta irá a buscar estas claves a Redis y obtendrá los resultados de los fragmentos de vídeo solicitados.

4.3.2 Persistencia de los datos

Otra de las características que implementamos es la posibilidad de persistir los videos e imágenes que los usuarios suben al sistema. De esta forma, el usuario cuenta con una galería de videos e imágenes ya procesados, junto con sus estímulos, y en caso de querer analizar los resultados nuevamente no es necesario volver a procesar el video o imagen, ya que los datos están almacenados.

Para lograr esto, nos fue necesario encontrar un servicio de almacenamiento capaz de guardar archivos de gran tamaño, como videos e imágenes y además, se necesitaba una base de datos para almacenar la información analizada de dichos archivos.

En la siguiente imagen se puede ver en detalle cómo los componentes API y Joiner interactúan con los servicios mencionados.

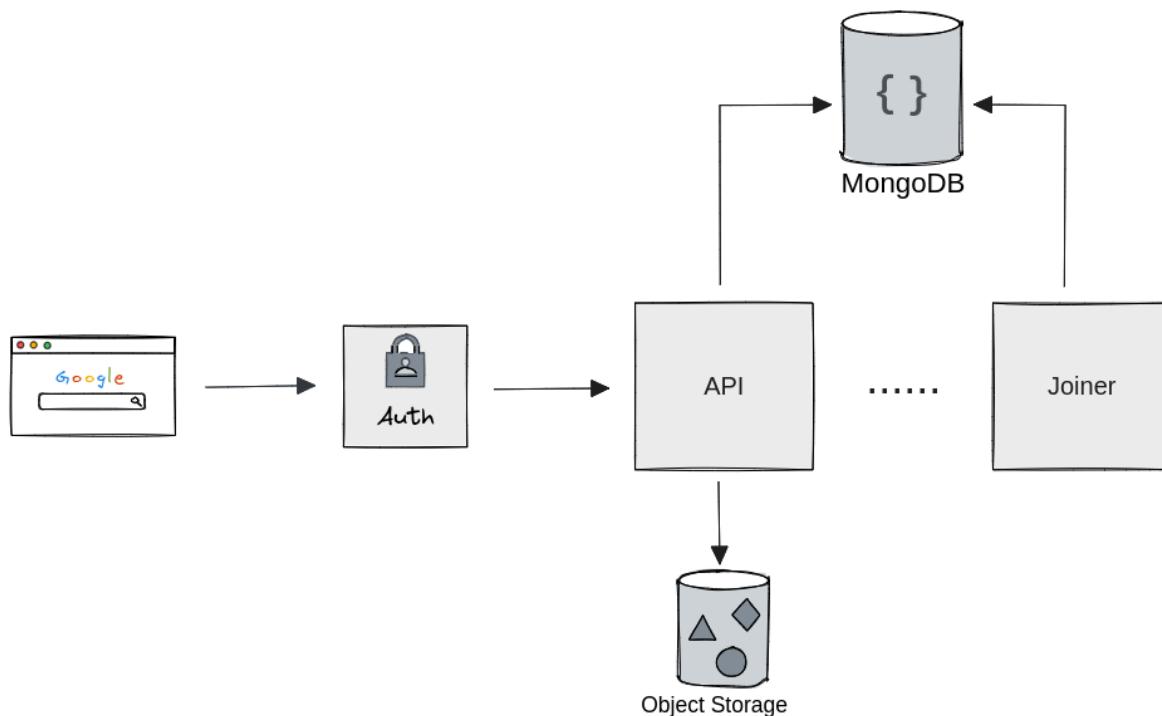


Figura 35: Arquitectura implementada para persistir información.

Para el almacenamiento de archivos multimedia, decidimos utilizar el Object Storage (OS) de Google Cloud Platform, el cual permite crear Buckets para almacenar ficheros de diferentes tipos de forma óptima y a un costo conveniente.

Dado que el Object Storage está diseñado específicamente para el uso que le damos, nos permitió también sacar provecho de otras funcionalidades que provee, por ejemplo el uso de URLs firmadas. Es decir, cuando un usuario quiere volver a ver un video o imagen previamente procesado, no necesitamos enviar todo el archivo mediante la API a la interfaz web, ya que el servicio nos brinda una URL que permite a la web cargar el video directamente desde la misma, facilitando el proceso.

Para el almacenamiento de los meta-datos elegimos usar una base de datos MongoDB, la cual brinda flexibilidad siendo capaz de manejar documentos no estructurados. También ofrece la posibilidad de en un futuro incorporar nuevos campos a los documentos, si así lo deseamos.

Dentro de MongoDB actualmente contamos con dos colecciones: `users` y `data`. La primera simplemente almacena los usuarios registrados, con un documento compuesto por el email y la contraseña encriptada. La segunda almacena un documento por cada archivo del usuario, dicho documento contiene toda la información necesaria para suplir las consultas de la interfaz web, como por ejemplo las *miniaturas* (thumbnails) para mostrar en la galería, nombre del archivo principal, nombre del estímulo, data pre-procesada si la hubiera, entre otros.

4.3.3 Comunicación Web-API

La web provee una interfaz para la carga del video a procesar y la carga del estímulo. Al momento de exponer los resultados, la web obtiene la información que debe mostrarse por segundo de video procesado. Con el fin de optimizar estas consultas, se decidió que la API devuelva un *batch* que tiene la información de hasta 10 segundos. En el siguiente diagrama se puede observar dicho flujo en un diagrama de secuencia.

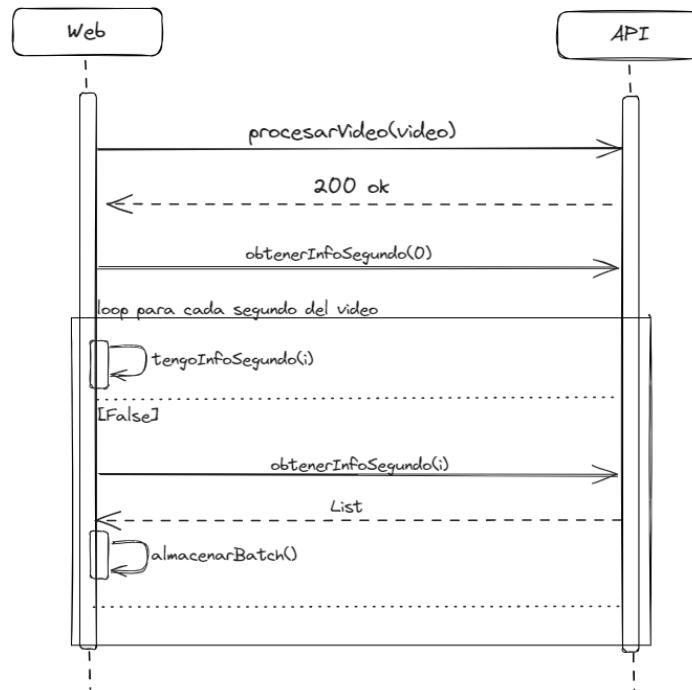


Figura 36: Diagrama de secuencia al procesar un video.

Cada vez que se debe actualizar los datos expuestos, primero se consulta si ya se cuenta con esos datos ya que estos podrían haberse devuelto dentro de un lote de respuesta anterior. En caso de que no se tengan los datos persistidos, se obtienen de la API. A su vez, la api podría devolver que ese segundo del video no fue procesado aun. En tal caso, se espera unos segundos para realizar un *retry*.

A medida que se va ejecutando el video, se van haciendo los llamados necesarios de la data futura para así poder lograr tener una respuesta lo más cercano a real time.

4.3.4 Procesamiento de Frames

¿Cuántos frames por segundo de video debemos procesar?

Si bien puede parecer algo sencillo, esta fue una de las decisiones claves del trabajo ya que debíamos buscar un equilibrio entre procesar menos frames, para lograr un mejor rendimiento

en cuanto tiempos de procesamiento, pero a la vez no debíamos perder precisión en el cálculo de nuestros resultados.

Para poder determinar hasta qué punto sería seguro reducir la cantidad de frames a procesar sin perder precisión, llevamos a cabo una prueba en la cual procesamos un mismo video dos veces, en la primera analizando todos los frames y en la segunda solo analizando un frame por segundo.

Luego calculamos la similitud entre los resultados obtenidos, tal como detallamos a continuación.

En primer lugar tenemos la variación de la excitación respecto al tiempo (frames). En azul para todos los frames y en rojo para un frame por segundo.

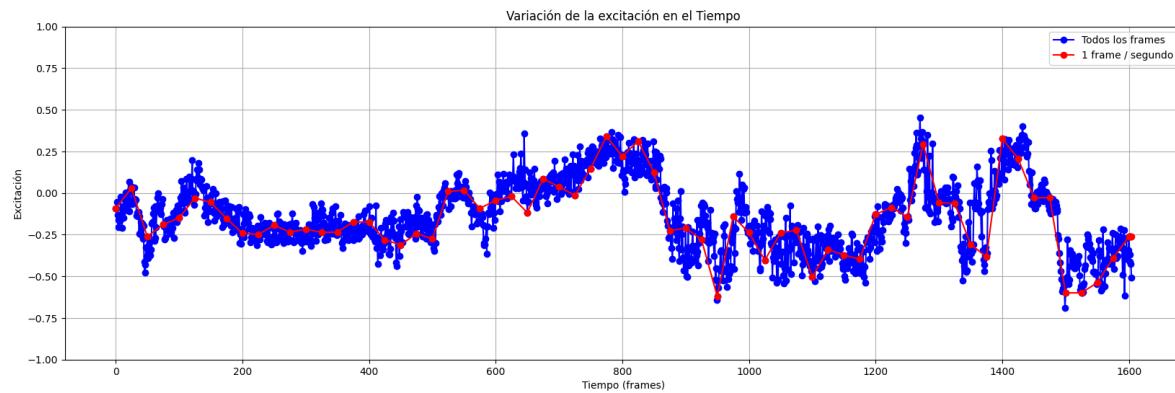


Figura 37: Variación de la excitación en el tiempo para un mismo video procesando todos los frames vs un frame por segundo.

Por otro lado, tenemos el mismo gráfico pero para la valencia:

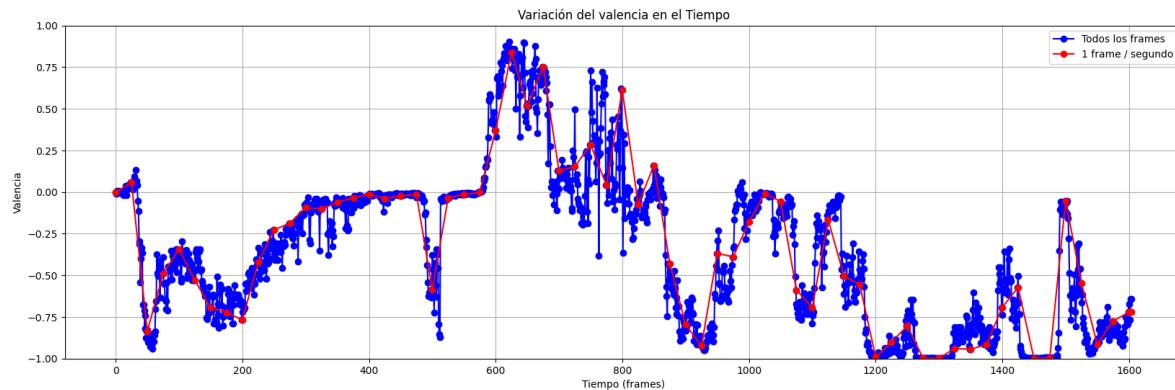


Figura 38: Variación de la valencia en el tiempo para un mismo video procesando todos los frames vs un frame por segundo.

A simple vista, podemos observar que tomando solo un frame por segundo los resultados parecen ser similares, en ningún momento vemos un desvío notorio con respecto al procesamiento en todos los frames.

Para poder tener un resultado más concreto, extendimos los resultado del procesamiento de un frame por segundo, obteniendo dos set de puntos de igual longitud, y calculamos la Correlación de Pearson [47] entre estos conjuntos, dando como resultado una similitud del 87.11% para la excitación y del 93.04% para la valencia.

De esta forma pudimos concluir que tomar solo un frame por segundo para nuestro análisis no afecta notoriamente la precisión de nuestros resultados pero si influye drásticamente, de forma positiva, en la velocidad de respuesta de nuestro sistema.

Tanto el notebook como la carpeta donde se encuentran los datos y el video utilizado se encuentran en el anexo.

4.3.4 Repositorios adicionales

Además de los componentes mencionados anteriormente, contamos con 3 repositorios adicionales:

- **Common:** expone un paquete de Python que puede ser importado por otras aplicaciones y contiene todo el código que es común en nuestros componentes, por ejemplo, la implementación de `Connection` para RabbitMQ que es utilizada tanto por la Api, Valence-processor y Joiner.
- **Rabbit:** Contiene lo necesario para poder crear una instancia de RabbitMQ local y así utilizar el sistema.
- **Storage:** Contiene lo necesario para poder instalar un emulador de GCP Object Storage e instancias de MongoDB y Redis. De esta manera podemos ejecutar todo el sistema de forma local, para realizar pruebas, sin necesidad de crear servicios cloud.

5. Metodología aplicada

Para el desarrollo del sistema, adoptamos un enfoque ágil basado en la metodología Scrum, con el objetivo de asegurar una gestión eficiente del proyecto y una entrega incremental del producto. A continuación, detallamos la estructura y los procesos implementados durante el desarrollo del proyecto.

5.1 Organización del Trabajo

El proyecto se dividió en sprints de dos semanas, facilitando una gestión ágil y flexible. Cada sprint comenzó con una reunión de planificación donde se definieron las tareas a realizar y se establecieron los objetivos específicos. Al finalizar cada sprint, se realizó una reunión de revisión en la cual cada miembro del equipo presentó los avances alcanzados, permitió identificar posibles impedimentos y planificar las próximas acciones.

Se establecieron reuniones semanales de seguimiento, en las que se discutían los progresos individuales y colectivos. Estas reuniones tuvieron un papel crucial en la coordinación del equipo y en la resolución de problemas emergentes. A su vez, si bien había tareas que desarrollamos de manera individual, algunas las desarrollamos de a pares, de manera que trabajamos en *programación en pareja (pair programming)*[48]. Además, tuvimos reuniones con el tutor que nos funcionaron como guías a la hora de investigar acerca de la computación afectiva y determinar el alcance del proyecto.

5.2 Herramienta de Trabajo

Para la gestión de tareas y el seguimiento del progreso, utilizamos la herramienta Jira. En esta herramienta creamos y asignamos las tareas correspondientes a cada sprint, permitiendo una visibilidad clara del estado del proyecto. El uso de esta Jira fue esencial para poder dividir la carga de trabajo, monitorear el progreso y ajustar el alcance en caso de ser necesario.

A su vez, implementamos un pipeline de CI/CD muy sencillo para automatizar el proceso de despliegue de los cambios. Esta práctica nos permitió detectar y corregir errores rápidamente, asegurando que las nuevas funcionalidades se integrarán sin problemas en el sistema existente y se desplegarán de manera eficiente.

6. Experimentación y/o validación

Para poder validar el sistema y los resultados obtenidos realizamos diferentes tipos de pruebas. En primer lugar, pruebas basales sobre los modelos de Machine Learning utilizados para poder asegurar que los resultados en las emociones detectadas y unidades de acción eran correctos. Luego, pruebas en las que comparamos los resultados de nuestro sistema (valencia y excitación) con resultados esperados ante cierto estímulo utilizando set de datos específicos para esto. Por último pruebas de campo en un contexto abierto.

6.1 Pruebas basales

Estas pruebas fueron realizadas en las etapas más tempranas del proyecto para poder validar que los modelos de Machine Learning elegidos para el sistema proporcionaban resultados coherentes.

Para las mismas, tomamos como referencia las imágenes de actores con las 6 expresiones faciales propuesta por Ekman, mencionado en la sección 2.2.2, figura 4, ya que conocíamos la emoción predominante en cada una de ellas. A continuación se detallan los resultados de las pruebas utilizando el modelo de Pyfeat para la detección de emociones básicas y el modelo de Openface para la detección de unidades de acción para cada una de estas emociones.

6.1.1 Asco

El análisis de la imagen en cuestión, que muestra a un actor representando la emoción de asco, utilizando nuestro sistema de detección de emociones basado en el modelo de Ekman, ha identificado predominantemente la emoción "asco" con una probabilidad de 0.975. La segunda emoción principal detectada fue "enojó", con una probabilidad de 0.021. Adicionalmente, al aplicar el modelo de Russel, la respuesta obtenida se ubica en el tercer cuadrante de su esquema. Estos resultados confirman que la emoción dominante reflejada en la imagen es el asco. El resultado de nuestro sistema se ve reflejado en la figura 39.

Herramienta para la evaluación de emociones en contextos abiertos

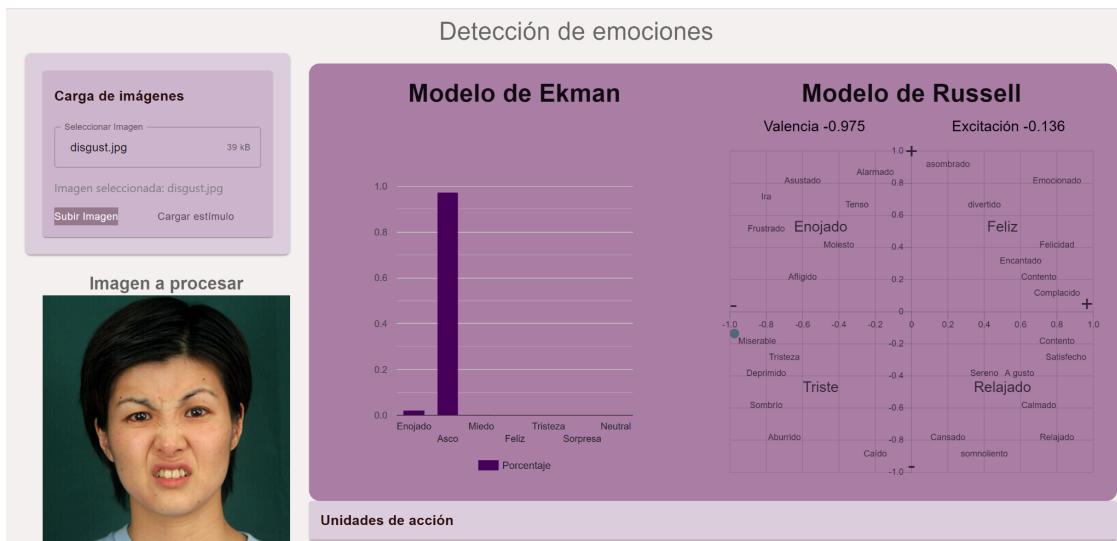


Figura 39: resultados del procesamiento de la imagen que representa ‘asco’

En la figura 40 se observan las unidades de acción detectadas en el rostro del actor que representa la emoción de “asco”. Las principales unidades de acción identificadas son AU9 (arrugador de nariz), AU7 (tensor del párpado) y AU10 (levantamiento del labio superior). Al comparar esta observación con la figura 3 de la sección 2.1.1 del Sistema de Codificación Facial (FACS), se concluye que las unidades de acción que deberían estar presentes o activas para representar dicha emoción son AU9 y AU10.

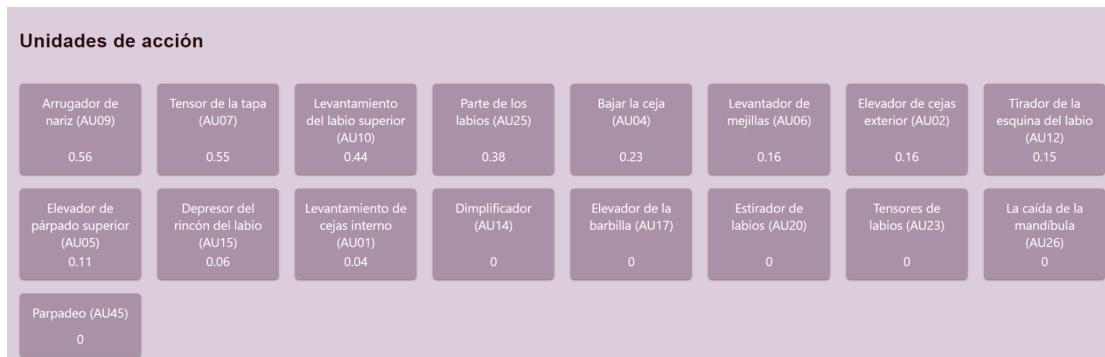


Figura 40: unidades de acción obtenidas por el sistema sobre la emoción ‘asco’.

6.1.2 Miedo

El análisis de la imagen en cuestión, que muestra a un actor representando la emoción de miedo, utilizando nuestro sistema de detección de emociones basado en el modelo de Ekman, ha identificado predominantemente la emoción "miedo" con una probabilidad de 0.992. La segunda emoción principal detectada fue "sorpresa", con una probabilidad de 0.006.

Herramienta para la evaluación de emociones en contextos abiertos

Adicionalmente, al aplicar el modelo de Russel, la respuesta obtenida se ubica entre el segundo y tercer cuadrante de su esquema. Estos resultados confirman que la emoción dominante reflejada en la imagen es el miedo. El resultado de nuestro sistema se ve reflejado en la figura 41.

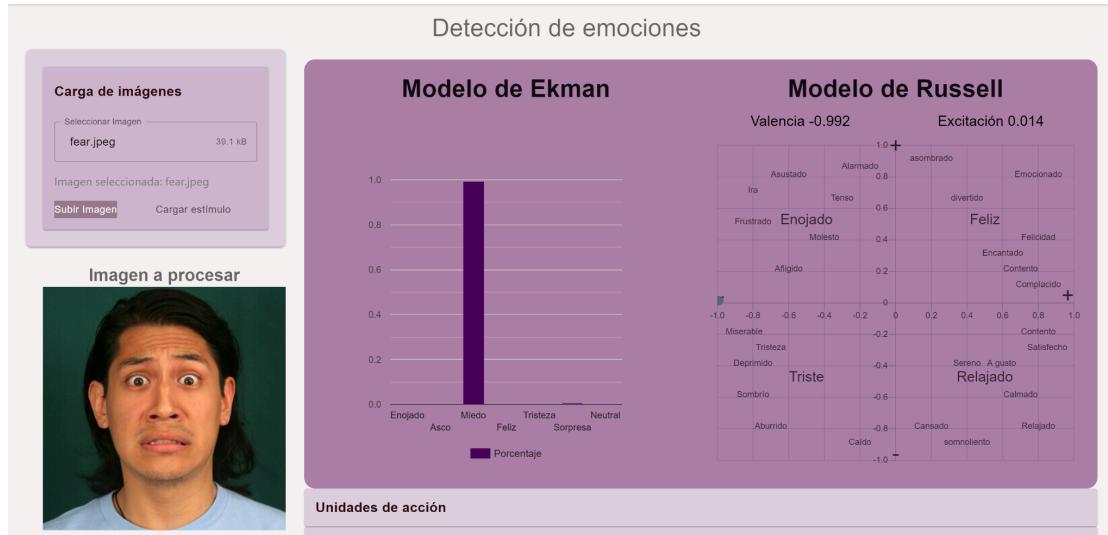


Figura 41: resultados del procesamiento de la imagen que representa ‘miedo’.

En la figura 42 se observan las unidades de acción detectadas en el rostro del actor que representa la emoción de “miedo”. Las principales unidades de acción identificadas son AU5 (Elevador del párpado superior), AU2 (Elevador de cejas externos), AU1 (Levantamiento de cejas interno), AU15 (Depresor del rincón del labio) y AU20 (Estirador de labios). Al comparar esta observación con la figura 3 de la sección 2.1.1 del Sistema de Codificación Facial (FACS), se concluye que las unidades de acción que deberían estar presentes o activas para representar dicha emoción son AU1, AU5, AU2, AU20, AU4 .

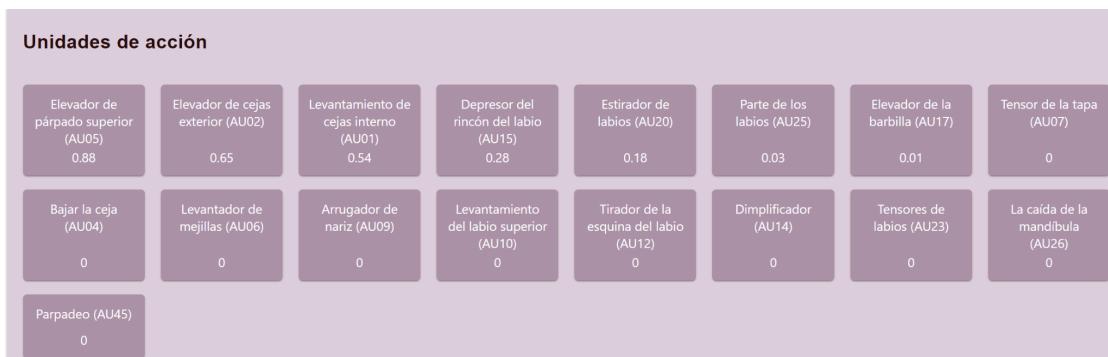


Figura 42: unidades de acción obtenidas por el sistema sobre la emoción ‘miedo’.

6.1.3 Felicidad

El análisis de la imagen en cuestión, que muestra a un actor representando la emoción de felicidad, utilizando nuestro sistema de detección de emociones basado en el modelo de Ekman, ha identificado predominantemente la emoción "feliz" con una probabilidad de 1. Adicionalmente, al aplicar el modelo de Russell, la respuesta obtenida se ubica entre el primer y cuarto cuadrante de su esquema. Estos resultados confirman que la emoción dominante reflejada en la imagen es la felicidad. El resultado de nuestro sistema se ve reflejado en la figura 43.

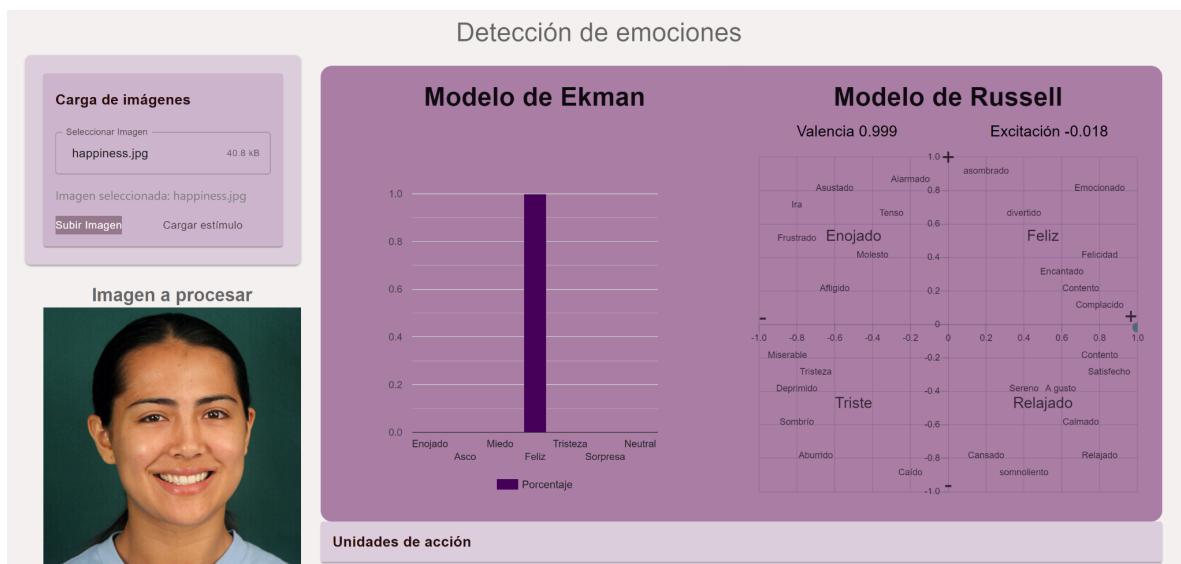


Figura 43: resultados del procesamiento de la imagen que representa ‘felicidad’.

En la figura 44 se observan las unidades de acción detectadas en el rostro del actor que representa la emoción de “felicidad”. Las principales unidades de acción identificadas son AU12 (Tirador de la esquina del labio), AU10 (levantamiento del labio superior), AU25 (parte de los labios) y AU6 (levantador de mejillas). Al comparar esta observación con la figura 3 de la sección 2.1.1 del Sistema de Codificación Facial (FACS), se concluye que las unidades de acción que deberían estar presentes o activas para representar dicha emoción son AU12 y AU6.

Herramienta para la evaluación de emociones en contextos abiertos

Unidades de acción							
Tirador de la esquina del labio (AU12) 0.71	Levantamiento del labio superior (AU10) 0.53	Parte de los labios (AU25) 0.42	Levantador de mejillas (AU06) 0.42	Tensor de la tapa (AU07) 0.37	Dimplificador (AU14) 0.31	Arrugador de nariz (AU09) 0.12	Levantamiento de cejas interna (AU01) 0
Elevador de cejas exterior (AU02) 0	Bajar la ceja (AU04) 0	Elevador de párpado superior (AU05) 0	Depresor del rincón del labio (AU15) 0	Elevador de la barbillia (AU17) 0	Estirador de labios (AU20) 0	Tensores de labios (AU23) 0	La caída de la mandíbula (AU26) 0
Parpadeo (AU45) 0							

Figura 44: unidades de acción obtenidas por el sistema sobre la emoción ‘felicidad’.

6.1.4 Tristeza

El análisis de la imagen en cuestión, que muestra a un actor representando la emoción de tristeza, utilizando nuestro sistema de detección de emociones basado en el modelo de Ekman, ha identificado predominantemente la emoción "tristeza" con una probabilidad de 0.837. La segunda emoción principal detectada fue "neutral", con una probabilidad de 0.122 y la tercera emoción principal detectada fue "miedo" con una probabilidad de 0.03. Adicionalmente, al aplicar el modelo de Russel, la respuesta obtenida se ubica en el tercer cuadrante de su esquema. Estos resultados confirman que la emoción dominante reflejada en la imagen es la tristeza. El resultado de nuestro sistema se ve reflejado en la figura 45.

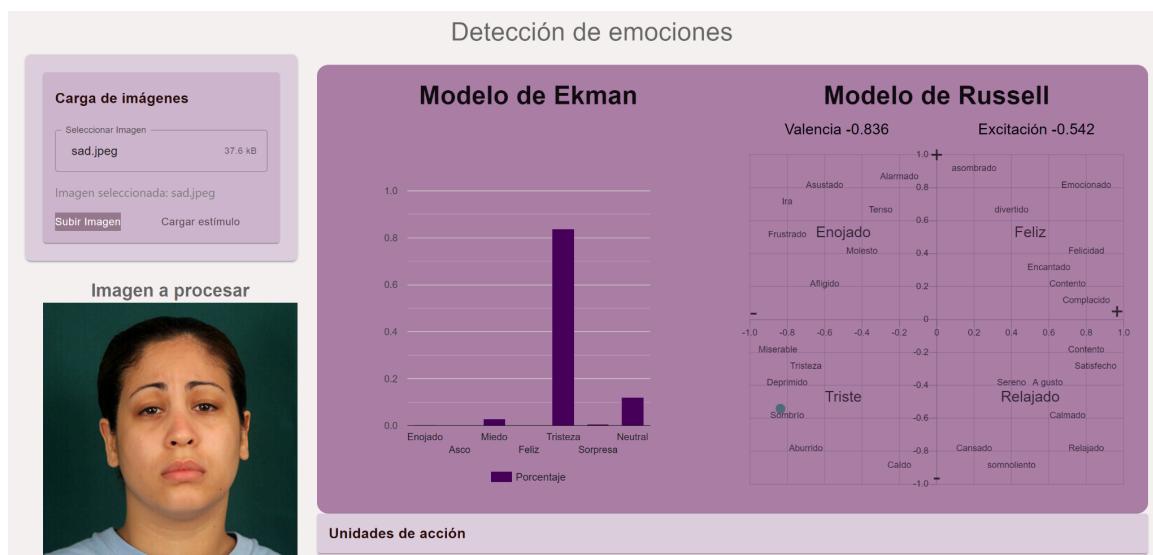


Figura 45: resultados del procesamiento de la imagen que representa ‘tristeza’.

En la figura 46 se observan las unidades de acción detectadas en el rostro del actor que representa la emoción de “tristeza”. Las principales unidades de acción identificadas son AU2 (Elevador de cejas externos), AU1 (Levantamiento de cejas interno) y AU15 (Depresor del

Herramienta para la evaluación de emociones en contextos abiertos

rincón del labio). Al comparar esta observación con la figura 3 de la sección 2.1.1 del Sistema de Codificación Facial (FACS), se concluye que las unidades de acción que deberían estar presentes o activas para representar dicha emoción son AU1, AU4 y AU15.

Unidades de acción							
Elevador de cejas exterior (AU02)	0.42	Levantamiento de cejas interno (AU01)	0.38	Depresor del rincón del labio (AU15)	0.14	Tensor de la tapa (AU07)	0.11
Parpadeo (AU45)	0.1	Bajar la ceja (AU04)	0.03	Elevador de párpado superior (AU05)	0	Elevador de mejillas (AU06)	0
Arrugador de nariz (AU09)	0	Levantamiento del labio superior (AU10)	0	Tirador de la esquina del labio (AU12)	0	Dimplificador (AU14)	0
Elevador de la barbilla (AU17)	0	Estirador de labios (AU20)	0	Tensores de labios (AU23)	0	Parte de los labios (AU25)	0
La caída de la mandíbula (AU26)	0						

Figura 46: unidades de acción obtenidas por el sistema sobre la emoción ‘tristeza’.

6.1.5 Sorpresa

El análisis de la imagen en cuestión, que muestra a un actor representando la emoción de sorpresa, utilizando nuestro sistema de detección de emociones basado en el modelo de Ekman, ha identificado predominantemente la emoción "sorpresa" con una probabilidad de 0.813. La segunda emoción principal detectada fue "miedo", con una probabilidad de 0.183. Adicionalmente, al aplicar el modelo de Russel, la respuesta obtenida se ubica entre el primer y cuarto cuadrante de su esquema. Estos resultados confirman que la emoción dominante reflejada en la imagen es sorpresa. El resultado de nuestro sistema se ve reflejado en la figura 47.

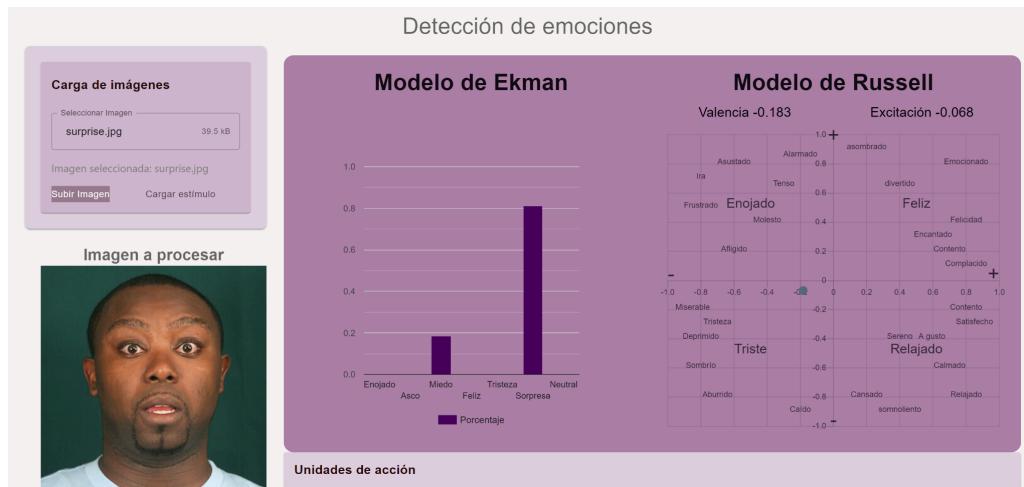


Figura 47: resultados del procesamiento de la imagen que representa ‘sorpresa’.

En la figura 48 se observan las unidades de acción detectadas en el rostro del actor que representa la emoción de “sorpresa”. Las principales unidades de acción identificadas son AU2 (Elevador de cejas exterior), AU5 (Elevador del párpado superior), AU1 (Levantamiento de cejas interno) y AU25 (Parte de los labios). Al comparar esta observación con la figura 3 de la sección 2.1.1 del Sistema de Codificación Facial (FACS), se concluye que las unidades de acción que deberían estar presentes o activas para representar dicha emoción son AU1, AU2, AU5 y AU26.

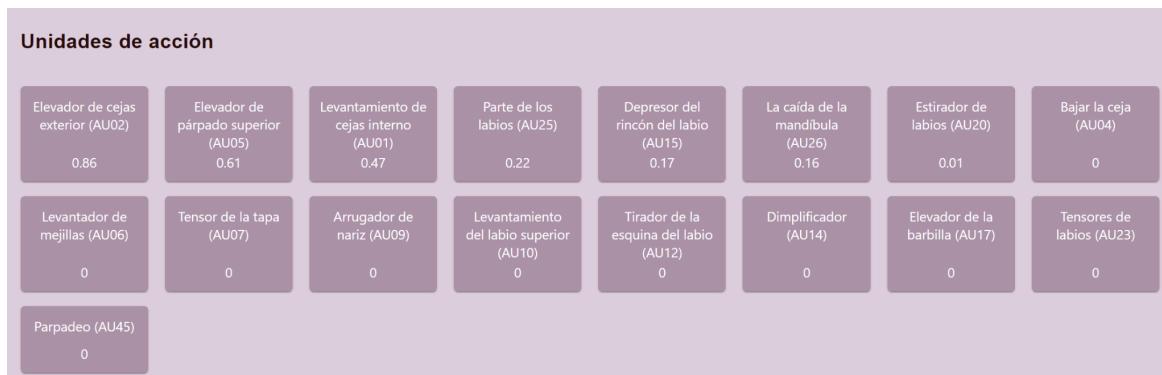


Figura 48: unidades de acción obtenidas por el sistema sobre la emoción ‘sorpresa’.

6.1.6 Enojo

El análisis de la imagen en cuestión, que muestra a un actor representando la emoción de enojo, utilizando nuestro sistema de detección de emociones basado en el modelo de Ekman, ha identificado predominantemente la emoción "enojado" con una probabilidad de 0.987. La segunda emoción principal detectada fue "neutral", con una probabilidad de 0.009. Si bien al aplicar el modelo de Russel la respuesta obtenida se ubica entre el tercer cuadrante de su esquema, afirmamos que los resultados confirman que la emoción dominante reflejada en la imagen es el enojo. El resultado de nuestro sistema se ve reflejado en la figura 49.

Herramienta para la evaluación de emociones en contextos abiertos

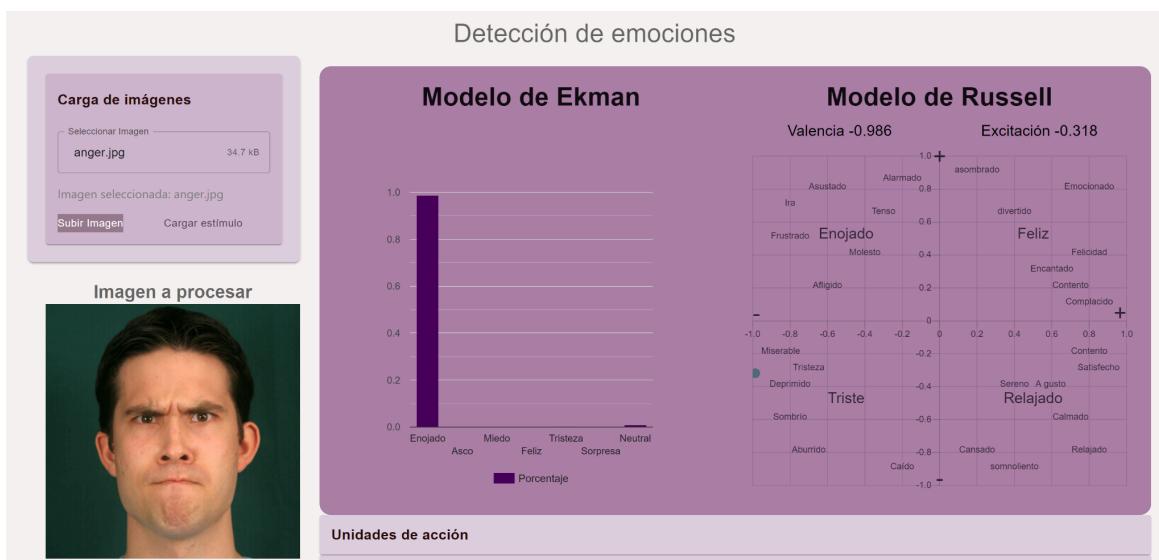


Figura 49: resultados del procesamiento de la imagen que representa ‘enojo’.

En la figura 50 se observan las unidades de acción detectadas en el rostro del actor que representa la emoción de “enojo”. Las principales unidades de acción identificadas son AU17 (Elevador de barbilla), AU4 (Bajar la ceja), AU26 (La caída de la mandíbula) y AU5 (Elevador del párpado superior). Al comparar esta observación con la figura 3 de la sección 2.1.1 del Sistema de Codificación Facial (FACS), se concluye que las unidades de acción que deberían estar presentes o activas para representar dicha emoción son AU5, AU4, AU17 y AU23.

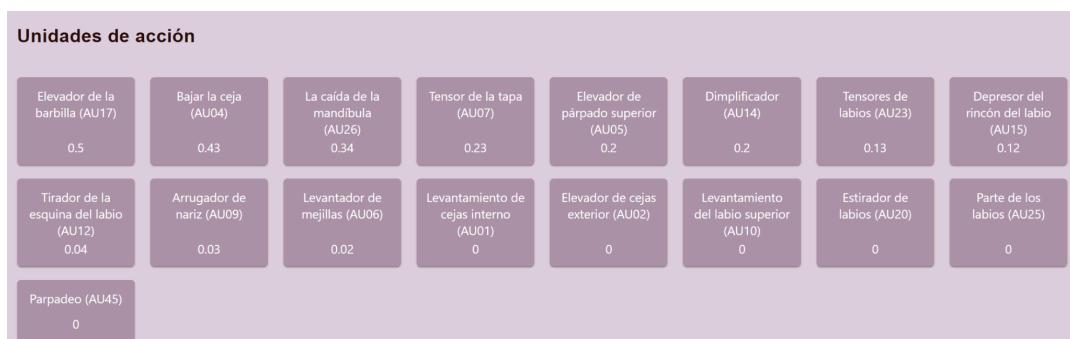


Figura 50 : unidades de acción obtenidas por el sistema sobre la emoción ‘enojo’.

En conclusión, se puede afirmar que el modelo empleado en el sistema tanto para el cálculo de las emociones básicas del modelo de Ekman como el modelo para la obtención de unidades de acción responde de manera efectiva frente a las imágenes de actores propuestas por Ekman. En general, el modelo detecta correctamente la emoción predominante en cada imagen con una probabilidad superior al 80%, demostrando su precisión y fiabilidad en el reconocimiento de

las emociones que los actores intentan representar. A su vez, el modelo detecta en general todas o prácticamente todas las principales unidades de acción que se encuentran asociadas a las emociones planteadas en la sección *2.1.1 del Sistema de Codificación Facial (FACS)*.

6.2 Pruebas de excitación y valencia comparadas con dataset DEVO

Para poder determinar si nuestra implementación del algoritmo de Noldus [22] para calcular valencia y excitación era correcta, utilizamos el set de datos DEVO [42] (Database of Emotional Videos), mencionado en la sección ‘*2.3 Estímulos y encuestas SAM*’.

Dado que este set de datos cuenta con valores promedio de excitación y valencia para los diferentes videos de estímulo, buscamos voluntarios a los que le grabamos el rostro mientras se reproducían estos videos como estímulos. Luego pudimos procesar estos videos para obtener nuestros resultados y compararlo con lo esperado. Los links a los videos se encuentran en el anexo.

La valencia y la excitación promedio proporcionadas por DEVO se encuentran en una escala de 1 a 9, donde para la valencia 1 corresponde a feliz y 9 a infeliz, y en cuanto a la excitación 1 significa excitado y 9 calmado. Por otro lado, nuestro sistema utiliza una escala de -1 a 1, donde una valencia de 1 indica felicidad y -1 indica infelicidad, y en términos de excitación, 1 representa excitación y -1 representa calma. Para poder realizar un análisis coherente, se reescalaron los datos de DEVO de manera lineal utilizando la siguiente fórmula:

$$Y = -\frac{1}{4}X + \frac{10}{8}$$

Para la desviación estándar se multiplicó por el siguiente factor de escala:

$$\text{factor} = \frac{\text{longitud del dominio del sistema}}{\text{longitud del dominio de DEVO}} = \frac{1-(-1)}{9-1} = \frac{1}{4}$$

Seleccionamos a una serie de individuos para participar en un conjunto de pruebas diseñadas para registrar sus reacciones frente a estímulos cuidadosamente escogidos por el equipo. Durante estas pruebas, se grabaron las reacciones faciales de los participantes y, posteriormente, se procesaron dichas grabaciones a través del sistema desarrollado por el equipo para realizar un análisis y evaluación exhaustivos de las emociones manifestadas.

A continuación, se detallan las pruebas realizadas para cada estímulo seleccionado, así como la información correspondiente a las personas involucradas en el proceso.

Previo a ver los resultados, debemos entender que las respuestas están abiertas a presentar ciertas diferencias más allá del desvío estándar debido a las distintas formas en las que fueron obtenidas. DEVO obtiene resultados a partir de realizar encuestas SAM a sus usuarios,

posterior a la visualización del video, mientras que nuestro sistema provee la respuesta instantánea a partir de expresiones faciales.

Por ejemplo en el video 21.6, la respuesta esperada se centra en mucha valencia. Pero el video es un simple fragmento relajante mostrando un paisaje de cascadas. Entendemos que al ver el video ninguna reacción se genera pero en un análisis posterior uno puede pensar que es un ambiente calmo y agradable produciendo un resultado posterior de valencia positivo. Por lo cual, lo más importante a considerar en los resultados es ver que ninguna de las respuestas son sustancialmente diferentes a la esperada por DEVO.

A su vez hay que comprender que DEVO es una muestra a un grupo reducido de un origen social similar y ajeno al propio de los voluntarios que nosotros utilizamos. Por lo que se puede esperar que ciertos videos tengan variaciones en lo esperado dependiendo del origen socio-cultural del visualizador. Como se podrá ver luego en el caso del video 96.1.

6.2.1 Video estímulo 28.1

En la tabla 3 se presentan los valores promedio esperados para la valencia y la excitación para el estímulo 28.1 cuya descripción es un bebé, proporcionados por el DEVO, junto con los valores promedio reescalados, que permiten realizar una comparación adecuada con el sistema.

Medida	DEVO media	Media reescalada	Devo desvío estándar	Desvío estándar reescalado
Valencia	1,92	0,77	1,62	0,41
Excitación	3,84	0,29	2,57	0,15

Tabla 3: Valencia y excitación para DEVO y reescalado para video 28.1

Por lo mencionado en la tabla 3, se esperaría que para las pruebas con los distintos individuos den como emoción predominante felicidad y que en el modelo de russell los valores promedios de excitación y valencia se encuentren entre el primer y cuarto cuadrante que representan las emociones de feliz y relajado.

6.2.1.1 Nazareno

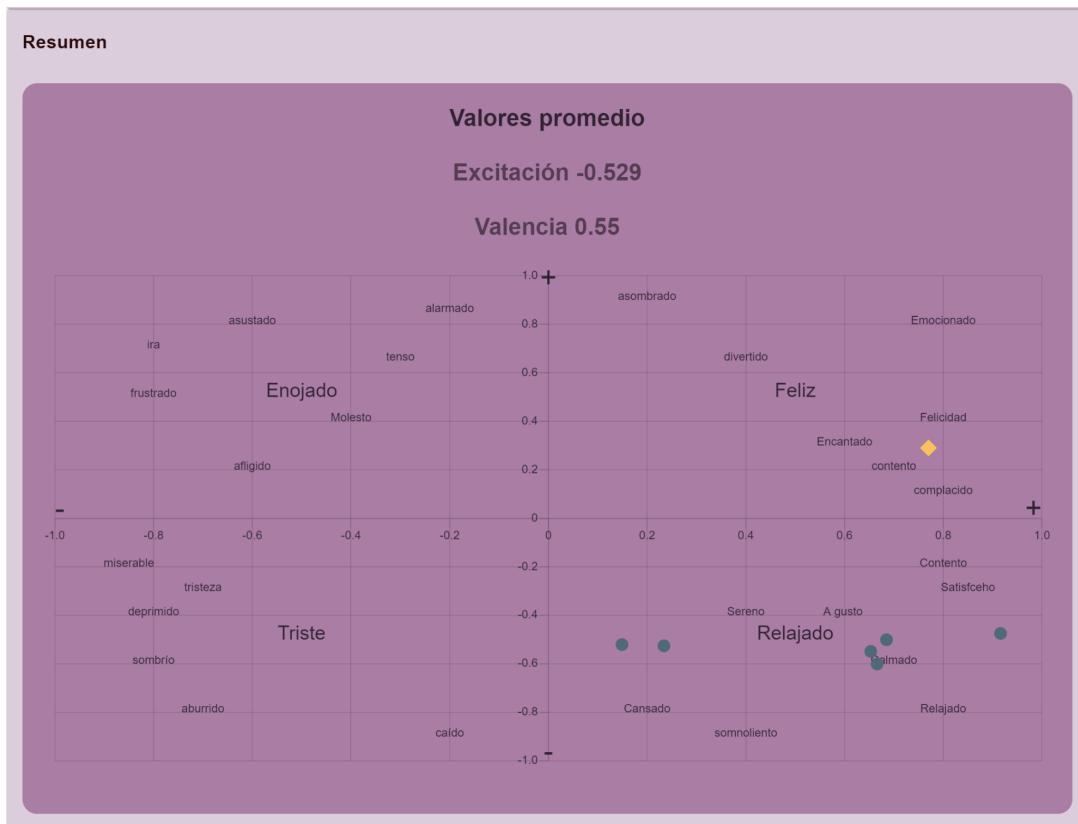


Figura 51: Resumen de puntos de valencia y excitación de nazareno frente al estímulo 28.1.

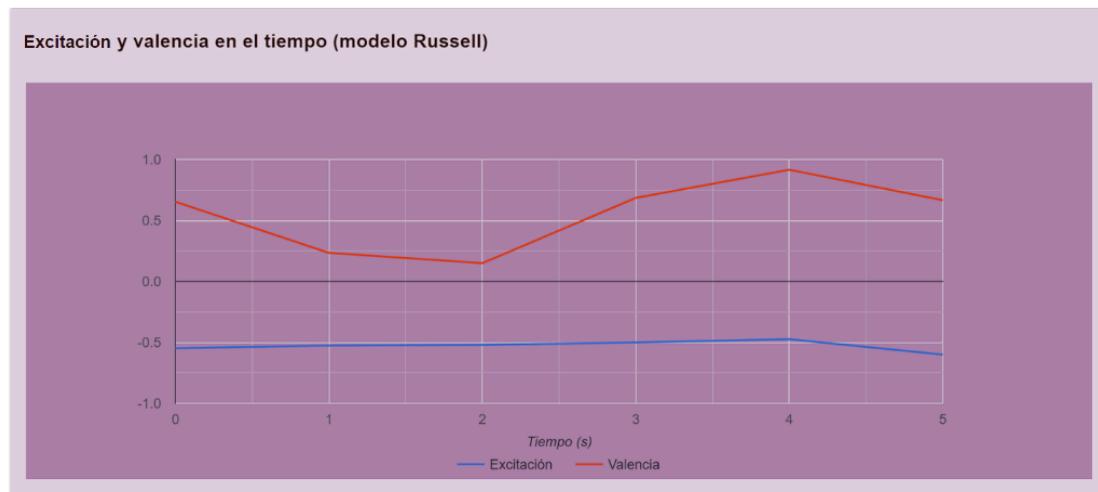


Figura 52: Gráfico en el tiempo de valencia y excitación de nazareno frente al estímulo 28.1.

A partir de las figuras 51 y 52, se observa que la valencia promedio calculada por el sistema es de 0.55. Este valor se encuentra dentro del rango esperado, considerando la desviación

Herramienta para la evaluación de emociones en contextos abiertos

estándar. En particular, aunque la valencia esperada es 0,77, su desviación estándar es 0,41, lo que implica que el valor obtenido de 0,55 se sitúa dentro de los límites de variabilidad aceptables según los parámetros definidos.

6.2.1.2 Valentina

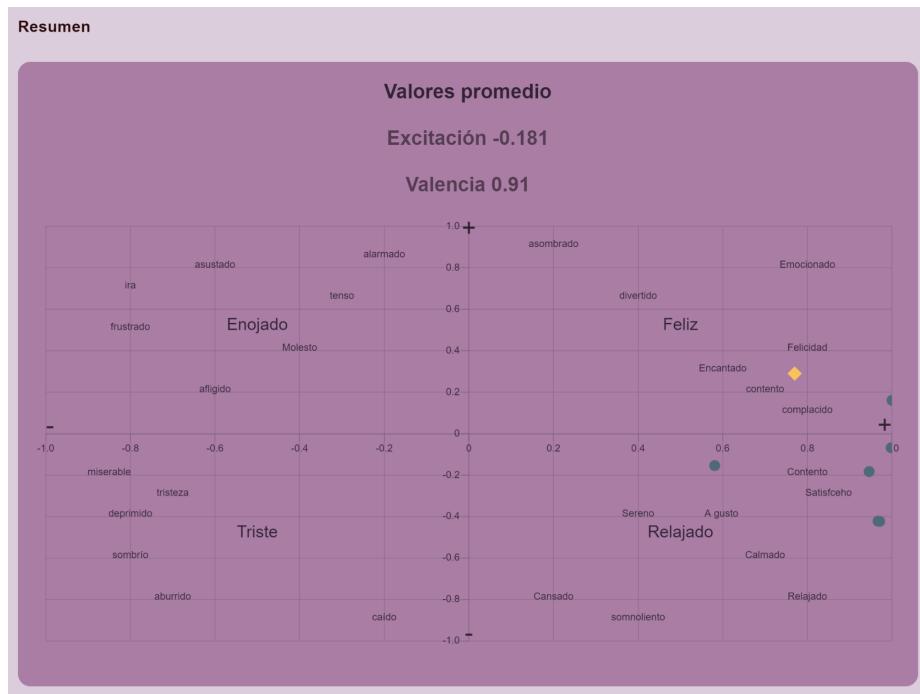


Figura 53: Resumen de puntos de valencia y excitación de Valentina frente al estímulo 28.1.

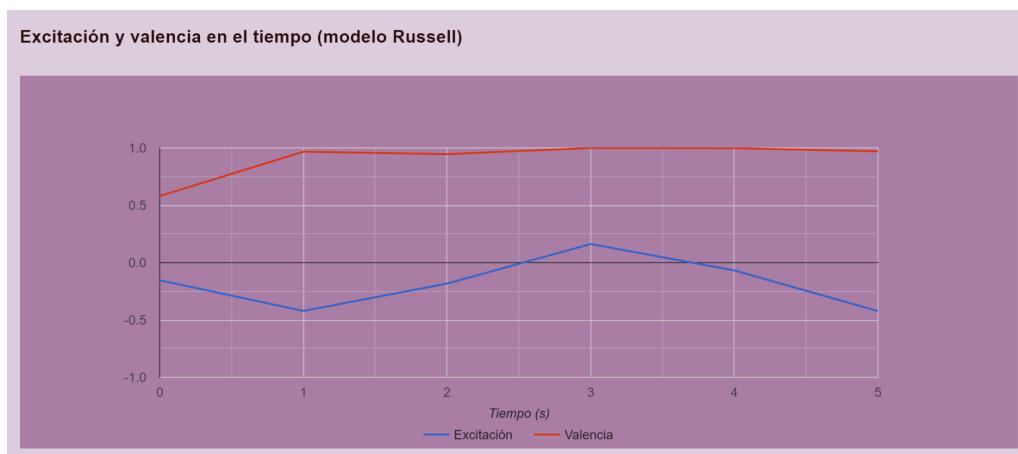


Figura 54: Gráfico en el tiempo de valencia y excitación de Valentina frente al estímulo 28.1.

Herramienta para la evaluación de emociones en contextos abiertos

Al igual que en la prueba anterior, se puede observar a partir de las figuras 53 y 54 que, aunque la valencia promedio obtenida fue bastante alta, se encuentra dentro de los parámetros esperados al considerar la desviación estándar. En particular, la valencia promedio calculada se ajusta adecuadamente a los valores previstos, demostrando la coherencia del sistema con las expectativas basadas en la desviación estándar.

Por otro lado, la excitación, aunque presenta un valor promedio de -0.18 y esta se encuentra un poco por fuera del rango de variabilidad definido por la desviación estándar, está acorde a los resultados esperados. Este resultado indica que tanto la valencia como la excitación, a pesar de sus valores específicos, se alinean con los parámetros esperados y confirman la precisión y fiabilidad del sistema en la evaluación de estas dimensiones emocionales.

Adicionalmente, se observa que los valores de valencia y excitación se mantienen dentro de los rangos esperados a lo largo del tiempo. Esta estabilidad temporal refuerza la validez de los resultados y la consistencia del sistema para medir las respuestas emocionales en diferentes momentos.

6.2.1.3 Deborah

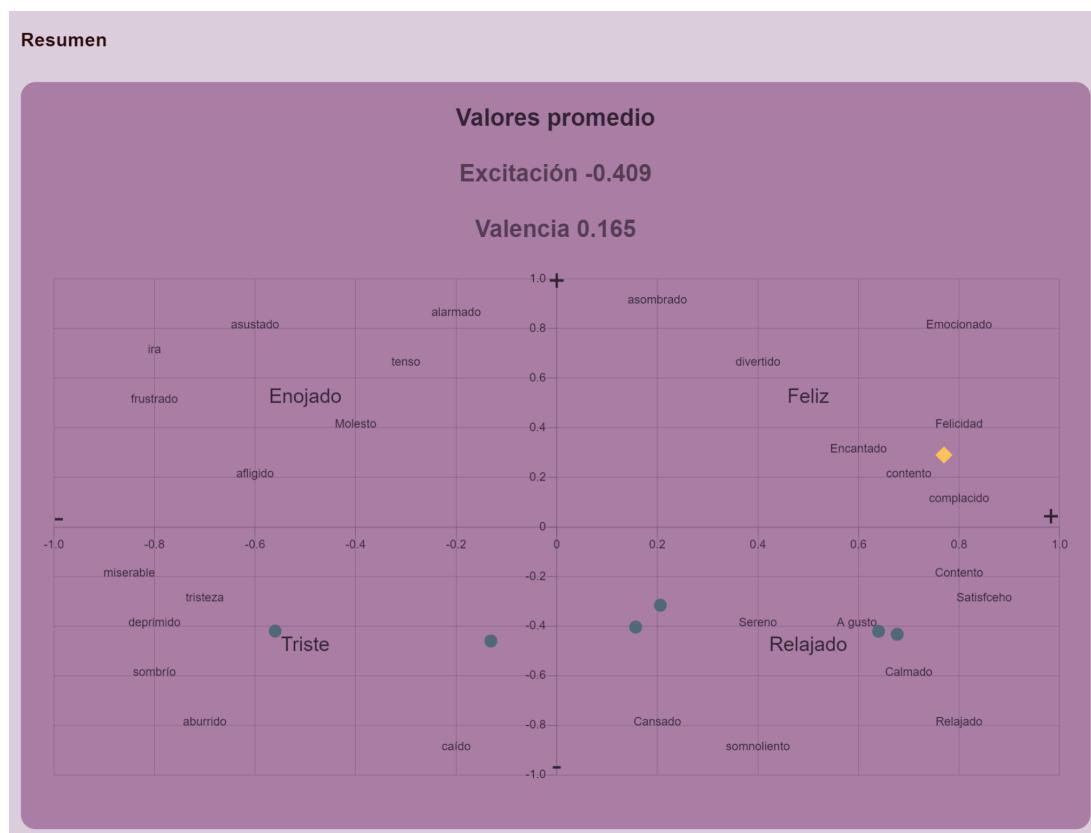


Figura 55: Resumen de puntos de valencia y excitación de Deborah frente al estímulo 28.1.



Figura 56: Gráfico en el tiempo de valencia y excitación de Deborah frente al estímulo 28.1.

De manera congruente con evaluaciones previas, los valores promedio de excitación y valencia se sitúan dentro de los parámetros esperados. No obstante, es relevante señalar la presencia de dos puntos particulares en los datos de valencia y excitación que inciden negativamente en el promedio de valencia. Estos puntos se localizan en el tercer cuadrante y corresponden al inicio y al final del video analizado. Es importante considerar que durante estos momentos, la persona pudo haber mantenido una actitud serena, sin reaccionar al inicio del video y al término del mismo.

Este análisis indica que los extremos emocionales observados al inicio y al final del estímulo podrían no reflejar las respuestas emocionales habituales que se manifestaron durante el transcurso del video.

6.2.1.4 Conclusión para el video 28.1

En conclusión, se observó que, en general, las personas reaccionaron de manera coherente con la emoción predominante durante todo el video, la cual fue la felicidad. Los valores de excitación y valencia se encontraron dentro de los cuadrantes esperados, confirmando su alineación con los rangos de valores previstos. Aunque se detectaron algunos casos atípicos (outliers) que ocasionaron ligeras modificaciones en los promedios esperados, estos no afectaron significativamente los resultados globales. En conjunto, los datos obtenidos demuestran que el sistema es efectivo y fiable para evaluar las respuestas emocionales, manteniéndose en consonancia con los parámetros establecidos.

6.2.2 Video 96.1

En la tabla 4 se presentan los valores promedio esperados para la valencia y la excitación para el estímulo 96.1 cuya descripción es un hombre vomitando, proporcionados por el DEVO, junto con los valores promedio reescalados, que permiten realizar una comparación adecuada con el sistema.

Medida	DEVO media	media reescalada	DEVO desvío estándar	Desvío estándar reescalado
Valencia	7,27	-0,56	1,91	0,19
Excitación	2,75	0,56	2,09	0,18

Tabla 4 : Valencia y excitación para devo y reescalado para video 96.1

Para esta prueba se esperaría que los individuos reaccionen de forma tal que su emoción principal sea el “asco” y que por lo tanto sus valores promedios de excitación y valencia se encuentren entre el segundo y tercer cuadrante que representan las emociones de enojo y tristeza.

6.2.2.1 Deborah

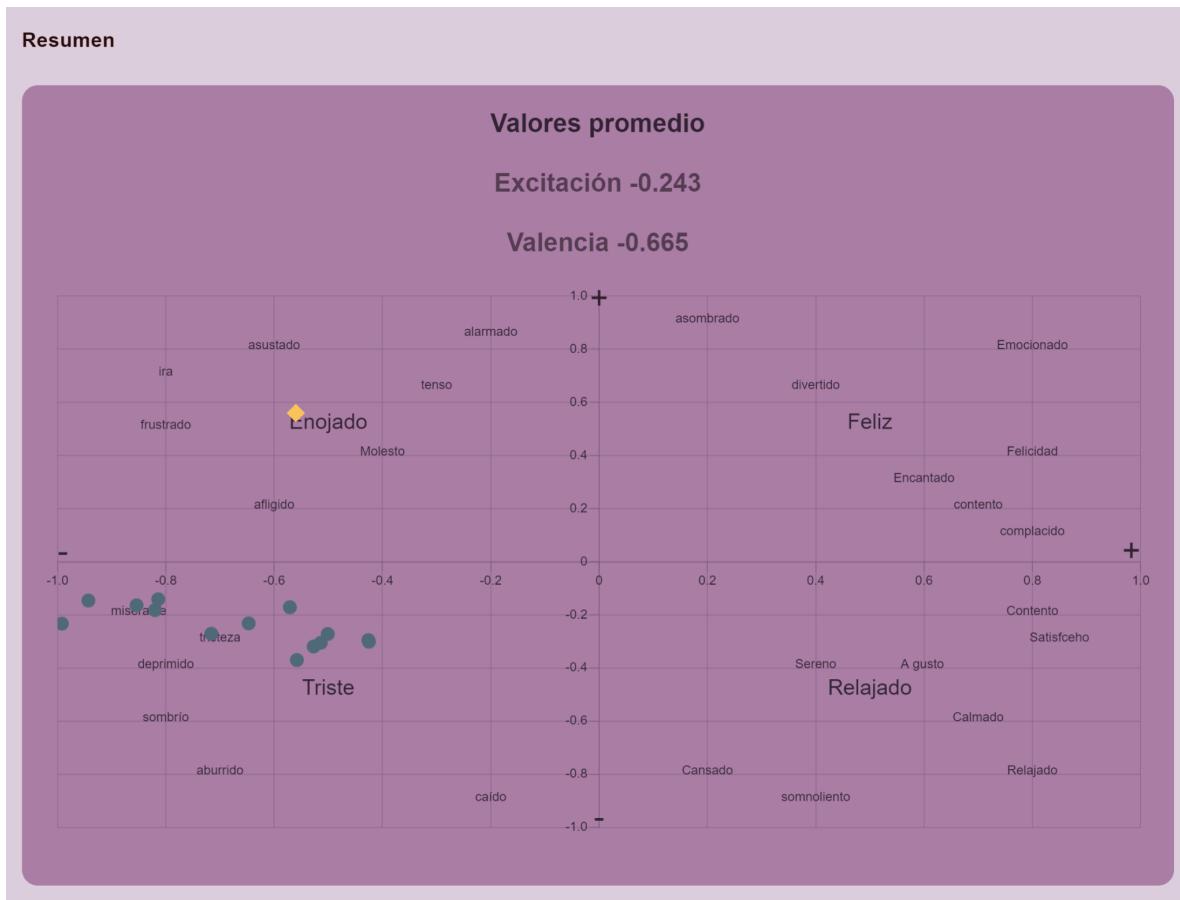


Figura 57: Resumen de puntos de valencia y excitación de Deborah frente al estímulo 96.1.

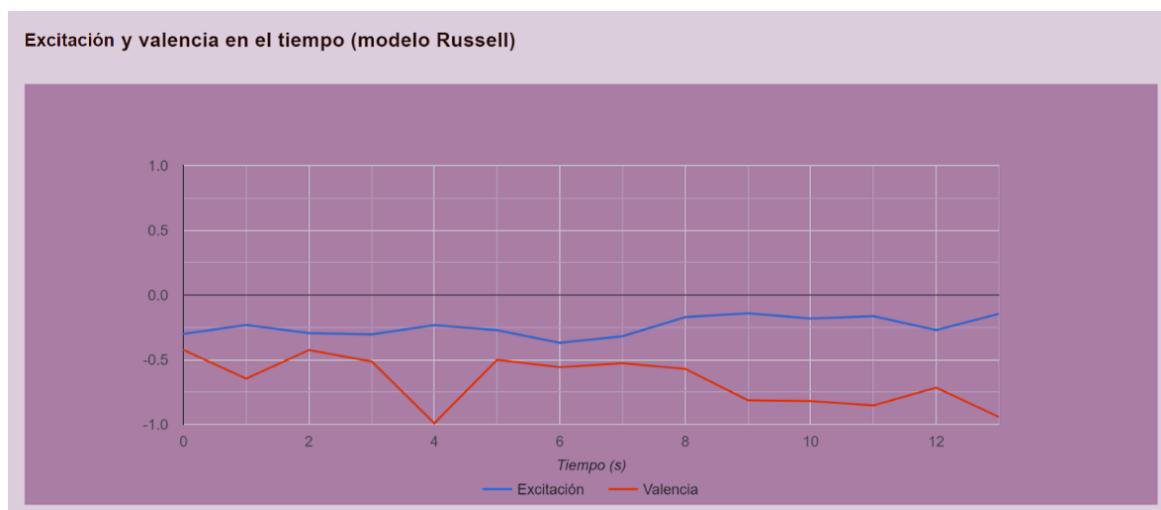


Figura 58: Gráfico en el tiempo de valencia y excitación de Deborah frente al estímulo 96.1.

Según los resultados de esta prueba, se observa que la valencia se aproximó al valor esperado de DEVO, mientras que la excitación mostró un desvío algo mayor. Sin embargo, el promedio de excitación se situó más cercano al eje de las abscisas que al mínimo valor de excitación.

Por otro lado, en general durante todo el video se detectó como principal emoción básica el “asco” en conjunto con “tristeza”

6.2.2.2 Conclusion Video 96.1

Durante esta prueba, se registró la participación de distintos individuos, quienes exhibieron variabilidad en sus respuestas emocionales. Algunos voluntarios mostraron gestos de gracia, mientras que otros manifestaron sorpresa. Destaca que Deborah fue la única persona que reaccionó conforme a las expectativas establecidas. Por este motivo, se llevó a cabo un análisis específico de los resultados obtenidos para ella. Los videos que muestran las reacciones de los demás participantes están disponibles en el drive adjunto en el anexo del presente informe.

6.2.3 Video 21.6

En la tabla 5 se presentan los valores promedio esperados para la valencia y la excitación para el estímulo 21.6 cuya descripción es una cascada, proporcionados por el DEVO, junto con los valores promedio reescalados, que permiten realizar una comparación adecuada con el sistema.

Medida	DEVO media	media reescalada	DEVO desvío estándar	Desvío estándar reescalado
Valencia	2,58	0,60	1,60	0,21
Excitación	6,02	-0,26	2,78	0,14

Tabla 5 : Valencia y excitación para dev0 y reescalado para video 21.6

Según lo indicado en la tabla 5, se esperaría que las pruebas realizadas con los distintos individuos muestran como emoción predominante el estado neutral y, como segunda emoción predominante, la felicidad. Asimismo, se anticipa que, en el modelo de Russell, los valores promedio de excitación y valencia se ubiquen en el cuarto cuadrante, que representa las emociones de relajación.

6.2.3.1 Roma

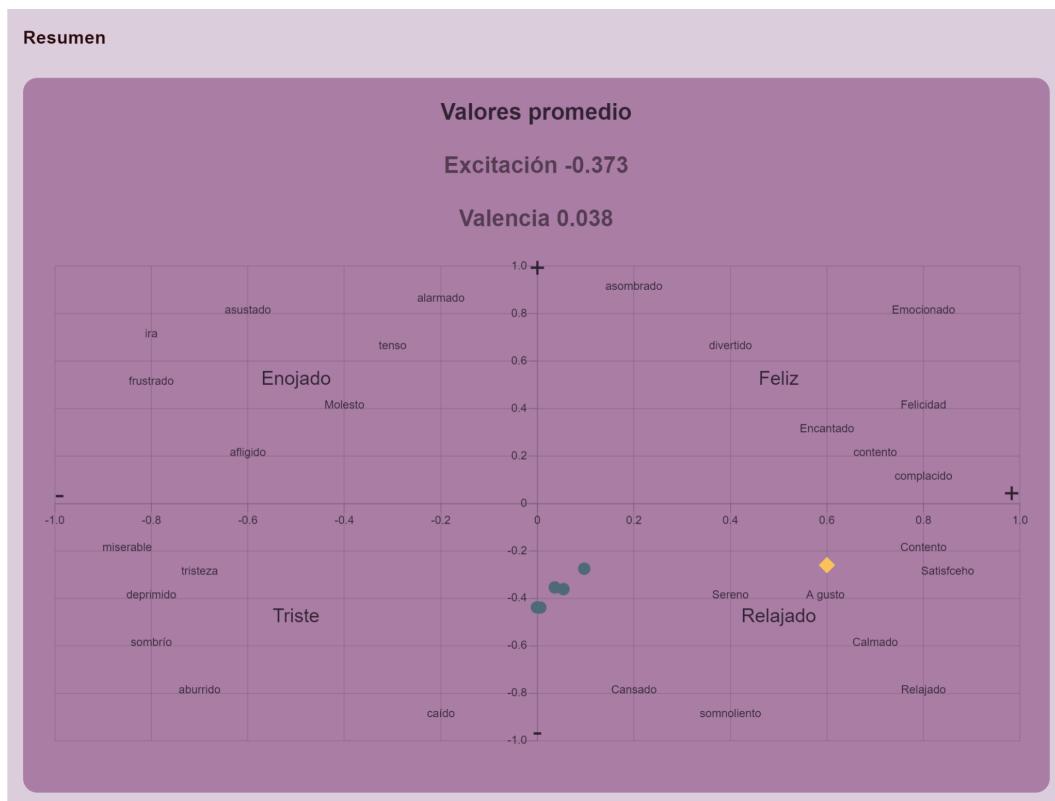


Figura 59: Resumen de puntos de valencia y excitación de Roma frente al estímulo 21.6

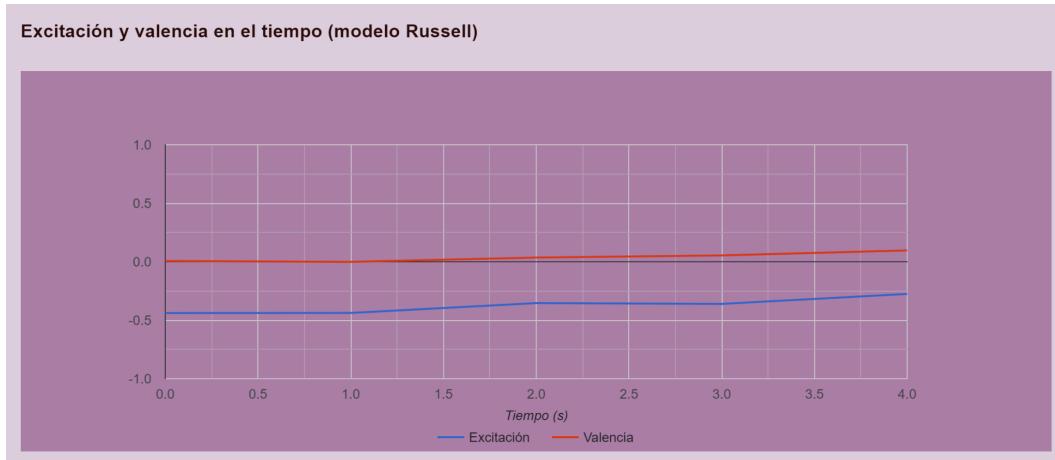


Figura 60: Gráfico en el tiempo de valencia y excitación de Roma frente al estímulo 21.6.

A partir de las figuras 59 y 60, se observa que el valor promedio de excitación obtenido por el sistema se asemeja al valor esperado proporcionado por DEVO. Por otro lado, aunque el

Herramienta para la evaluación de emociones en contextos abiertos

valor promedio de valencia obtenido se encuentra algo alejado del valor esperado, este se mantiene dentro del rango aceptable considerando la desviación estándar. Además, los gráficos demuestran que tanto la valencia como la excitación se mantienen constantes a lo largo del tiempo.

La diferencia en la valencia se debe a que el modelo, en general, detectaba la emoción "neutral" como predominante y "felicidad" como la segunda emoción más frecuente. Esta detección influye en los valores promedios obtenidos, ajustándose dentro de los rangos esperados según las variaciones naturales observadas.

Asimismo, se puede determinar que tanto el cuadrante esperado como el cuadrante calculado coinciden, predominando el cuarto cuadrante, que corresponde a la emoción de relajación. Este resultado reafirma la eficacia del sistema para evaluar y categorizar las respuestas emocionales dentro de los parámetros establecidos por el modelo de Russell.

En conclusión, los datos analizados confirman que el sistema de medición de emociones proporciona resultados coherentes y confiables, manteniendo los valores de excitación y valencia dentro de los rangos esperados y reflejando adecuadamente las emociones predominantes detectadas durante las pruebas.

6.2.3.2 Valentina



Figura 61: Resumen de puntos de valencia y excitación de Valentina frente al estímulo 21.6.

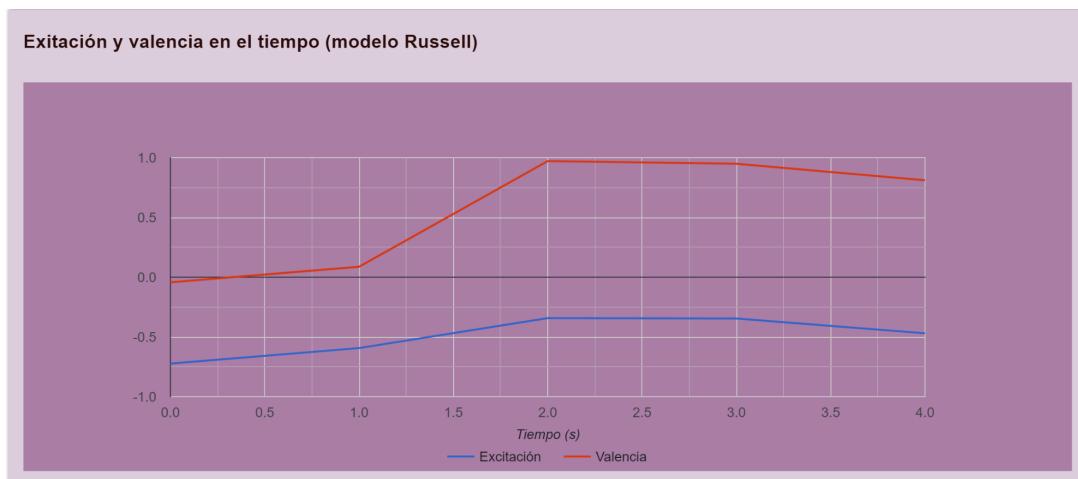


Figura 62: Gráfico en el tiempo de valencia y excitación de Valentina frente al estímulo 21.6.

Durante este análisis, se obtuvieron resultados similares a los del participante anterior (Roma). A partir de las figuras 61 y 62, se puede observar que la excitación promedio mostró un valor muy próximo al esperado, al igual que la valencia. Los valores esperados coincidieron en términos de cuadrantes, específicamente en el tercer cuadrante del modelo de Russel, que representa las emociones de relajación.

6.2.3.3 Conclusion Video 21.6

En términos generales, los resultados obtenidos para ambos videos muestran que el modelo de Russell identificó consistentemente el cuarto cuadrante esperado, correspondiente a la emoción de relajación. Además, se observó que ambos videos presentaron una excitación promedio similar a la esperada. Sin embargo, en el primer caso se registró una desviación significativa respecto al valor esperado de valencia mientras que en el segundo se obtuvo una valencia prácticamente igual a la esperada. A pesar de esta variación, los valores de valencia calculados se encuentran dentro del rango aceptable cuando se considera la desviación estándar.

Estos hallazgos indican que el sistema de medición de emociones utilizado proporciona resultados coherentes y confiables en la categorización de las respuestas emocionales, manteniéndose consistentemente alineado con los parámetros esperados para la excitación y el cuadrante emocional identificado. La ligera discrepancia en los valores de valencia sugiere posibles variaciones individuales en la interpretación de las emociones, aunque estos resultados no comprometen la validez global del sistema para analizar y comparar las respuestas emocionales en los videos evaluados.

6.2.4 Pruebas de valencia y excitación - conclusiones

A partir de las pruebas realizadas, hemos logrado obtener un análisis más preciso y fundamentado.

Es fundamental destacar que, si bien los videos de DEVO muestran valores esperados en términos de valencia y excitación junto con su desviación estándar, las reacciones de los individuos están inherentemente influenciadas por factores subjetivos, incluyendo el contexto en el cual fueron grabadas y las experiencias personales de los participantes. Por consiguiente, es crucial enfatizar que los resultados no son absolutos, sino que están sujetos a subjetividades.

Como conclusión general, se observó que ninguna reacción mostró emociones opuestas o significativamente alejadas de las esperadas. Se destaca particularmente el cálculo de la valencia, el cual demostró ser más preciso en comparación con la excitación, la cual en muchos casos mostró una desviación considerable respecto al valor esperado.

6.3 Prueba de campo

Por último realizamos una prueba de campo para analizar cómo se comporta el sistema en un contexto abierto, para el cual no contamos con los valores de excitación y valencia promedios del estímulo, y así poder comparar los resultados obtenidos con lo “esperado”. El video analizado consiste en una persona que se graba mientras juega (*gameplay*) un videojuego de terror [50].

Dado el contexto, en este caso un videojuego de terror, esperamos que la reacción del jugador ante la aparición de un estímulo sea de sorpresa, enojo y/o miedo.

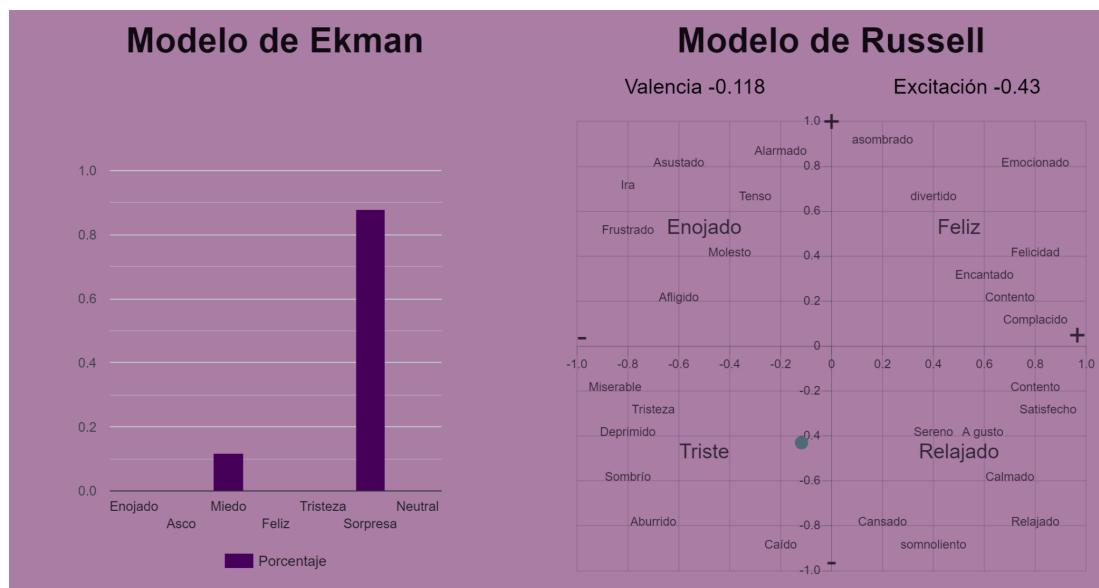


Figura 63: resultados de la prueba de campo en el minuto 0:02.

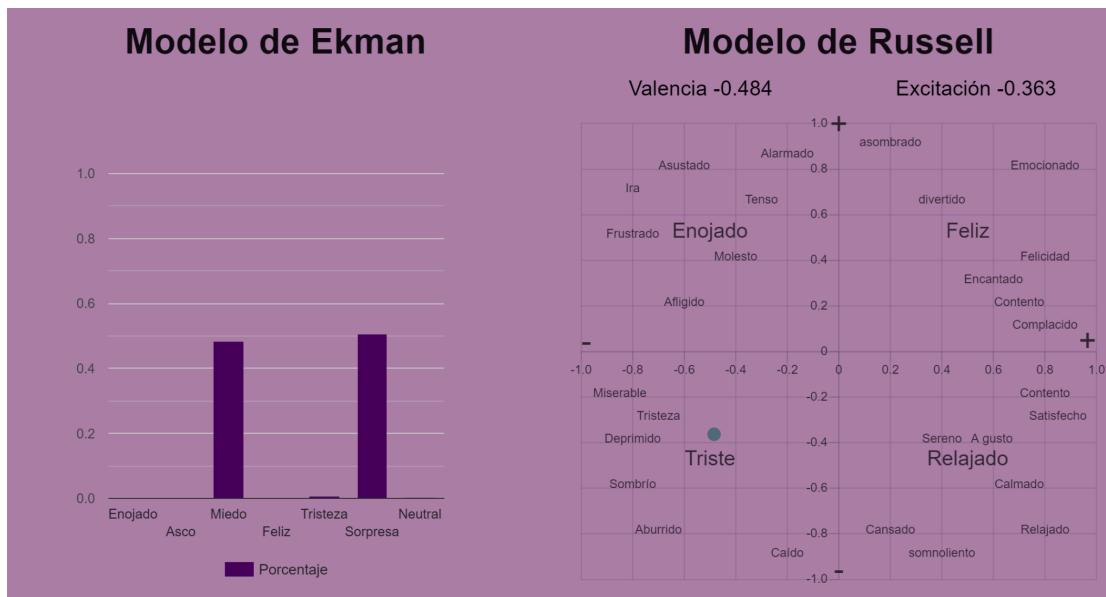


Figura 64: resultados de la prueba de campo en el minuto 0:12.

En la figura 63 y 64 se observa que el jugador exhibió emociones principales de sorpresa, miedo y tristeza tanto en el minuto 0.02 como en el minuto 0.12. Los valores de valencia registrados oscilan entre -0.1 y -0.5, mientras que los valores de excitación se sitúan entre -0.3 y -0.6. Estos resultados posicionan predominantemente al jugador en el tercer cuadrante del modelo de Russell, caracterizado por la emoción de tristeza.

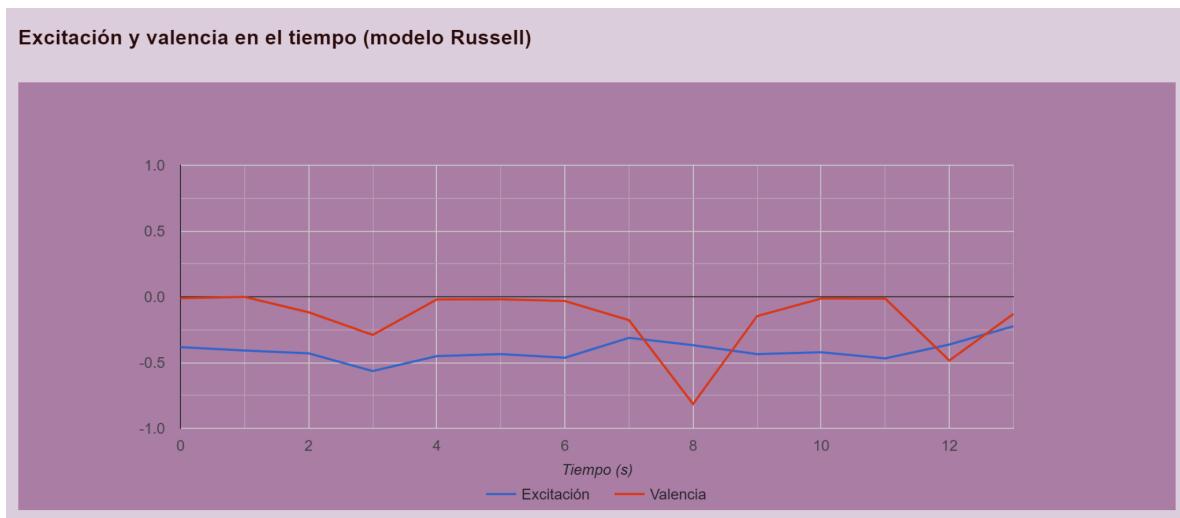


Figura 65: variación de la excitación y de la valencia de la prueba a lo largo del video.

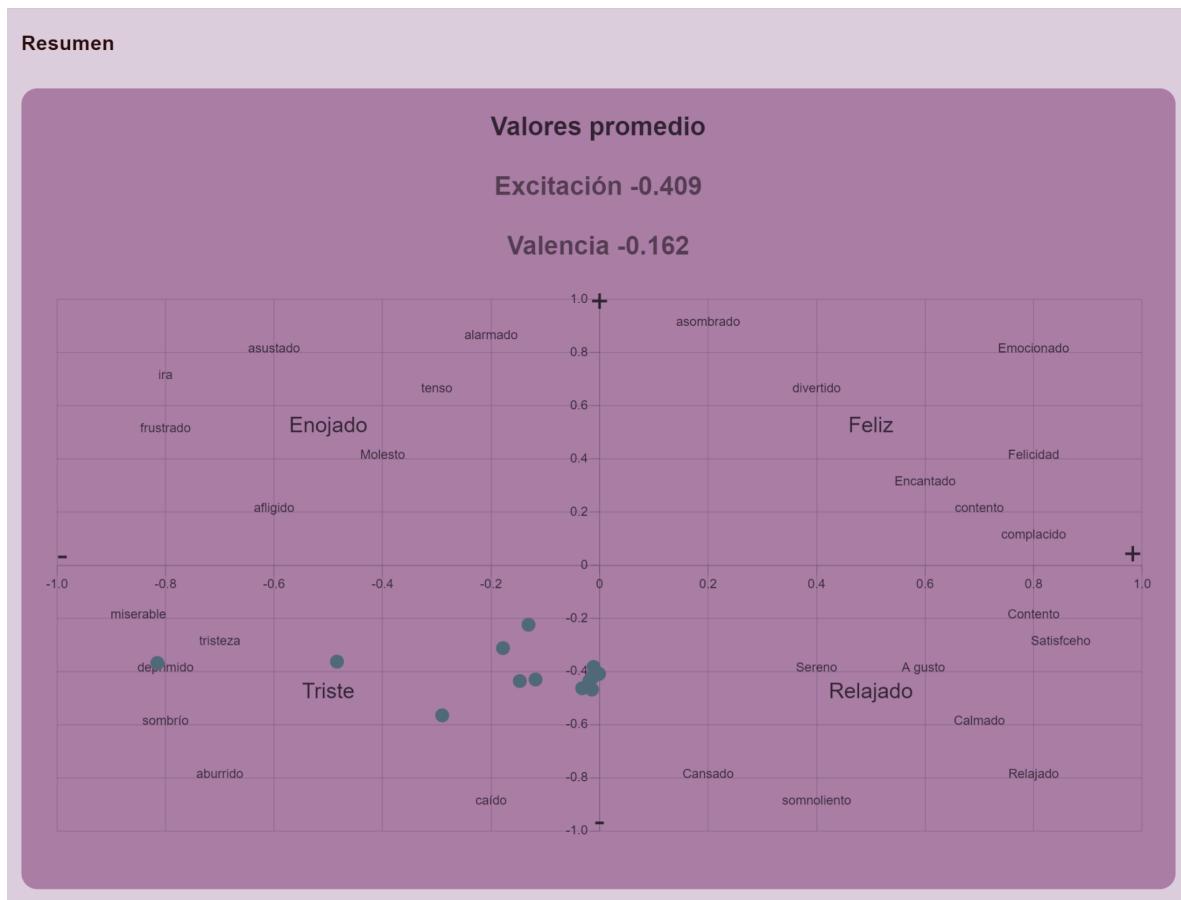


Figura 66: Resumen de puntos de valencia y excitación de la prueba de campo.

La nube de puntos representada en la figura 66 indica que durante la duración del video, todos los puntos se encuentran principalmente en el tercer cuadrante del modelo de Russell. Este patrón es consistente con los datos mostrados en la figura 65, donde se observa que los valores de valencia y excitación muestran poca variación a lo largo del tiempo, excepto en el segundo 8, donde la valencia experimenta una notable caída hasta alcanzar un valor de -0.8.

Los valores promedio calculados para excitación y valencia son -0.41 y -0.16, respectivamente. Estos resultados eran previsibles dado que se trata de un videojuego de miedo, sugiriendo que la persona experimentó predominantemente esta emoción durante la interacción con el juego.

En conclusión, los análisis realizados indican que el sistema de medición de emociones está funcionando conforme a lo esperado, en línea con las emociones anticipadas para un contexto de videojuego de terror. La consistencia de los datos en el tercer cuadrante del modelo de Russell y la coherencia de los valores promedio obtenidos refuerzan la validez y precisión del sistema para evaluar y categorizar las respuestas emocionales observadas.

6.4 Validación con Morphcast

Morphcast es una herramienta avanzada de análisis de video que emplea inteligencia artificial para interpretar las emociones y comportamientos de los individuos en tiempo real. Esta tecnología se utiliza en diversos ámbitos, incluyendo marketing, educación y seguridad, para proporcionar insights valiosos sobre la respuesta emocional y el comportamiento de los usuarios. Morphcast permite una evaluación detallada de las expresiones faciales, los movimientos y otros indicadores emocionales, brindando un análisis profundo y preciso.

En este contexto, decidimos realizar un análisis utilizando Morphcast con el propósito de comparar su precisión con nuestro propio sistema de análisis de video. Para ello, se emplearán distintos videos utilizados en nuestros análisis anteriores.

El objetivo de esta comparación es evaluar la eficacia y precisión de nuestro sistema en el análisis de emociones y comportamientos en comparación con MorphCast. Al utilizar los mismos videos, se asegura que las condiciones de prueba son idénticas, permitiendo una evaluación justa y objetiva de ambas herramientas. Se analizarán aspectos como la precisión en la detección de emociones, detección de valencia y excitación y la capacidad de interpretación de datos.

6.4.1 Análisis para prueba de campo

Para llevar a cabo la comparación, se procesó el video utilizado como prueba de campo, el cual consiste en una persona jugando a un videojuego de terror, a través de la herramienta Morphcast. Este video fue seleccionado debido a la intensa y variada gama de emociones que se pueden observar mientras el jugador reacciona a los elementos de sorpresa y miedo presentes en el juego.

En la figura 67 podemos observar que los valores arrojados por MorphCast son muy similares a los obtenidos por nuestro sistema.

Para la muestra del minuto 0:02 de video vemos que la emoción predominante es ‘Sorpresa’ con una probabilidad del 0.77, seguida de ‘Miedo’ con una probabilidad de 0.1. Comparando con los resultados de nuestro sistema, de la figura 63, podemos observar que los resultados son muy similares ya que también obtuvimos esas emociones predominantes con probabilidades muy parecidas.

Por otro lado, los valores de excitación y valencia arrojaron -0.46 y -0.25, respectivamente. Esto también tiene mucha correlación con lo obtenido en nuestro experimento, donde los valores para el minuto 0:02 eran -0.43 para la excitación y -0.118 para la valencia, como también se puede ver en la figura 63.

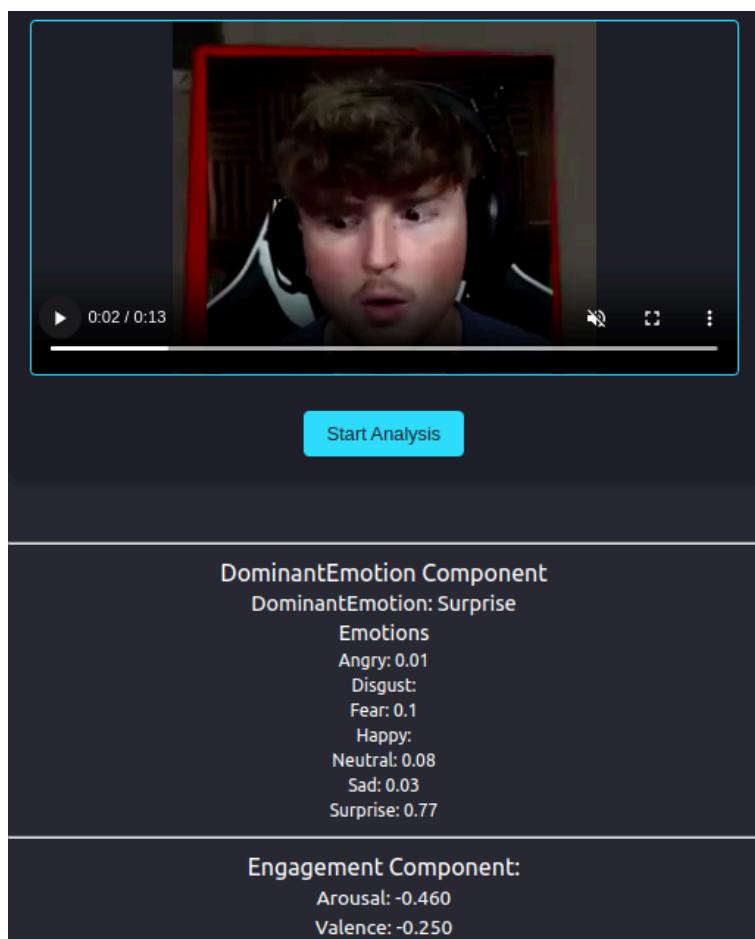


Figura 67: Prueba de campo realizada con MorphCast, muestra del minuto 0:02.

Por otro lado también podemos observar como los valores promedio de excitación y valencia al final de los experimentos son similares. En donde nuestro sistema calcula -0.409 para la excitación y -0.162 para la valencia, mientras que MorphCast arroja -0.468 y -0.267 respectivamente. Estos valores se pueden ver en la siguiente figura.

Mean values:
Arousal mean: -0.468
Valence mean: -0.267

Figura 68: Valores promedio de excitación y valencia arrojados por MorphCast para la prueba de campo.

Herramienta para la evaluación de emociones en contextos abiertos

Lo cual también se condice con la figura 69 en la cual podemos ver que también detecta emociones en el tercer cuadrante, como lo hace nuestro sistema en la figura 63.

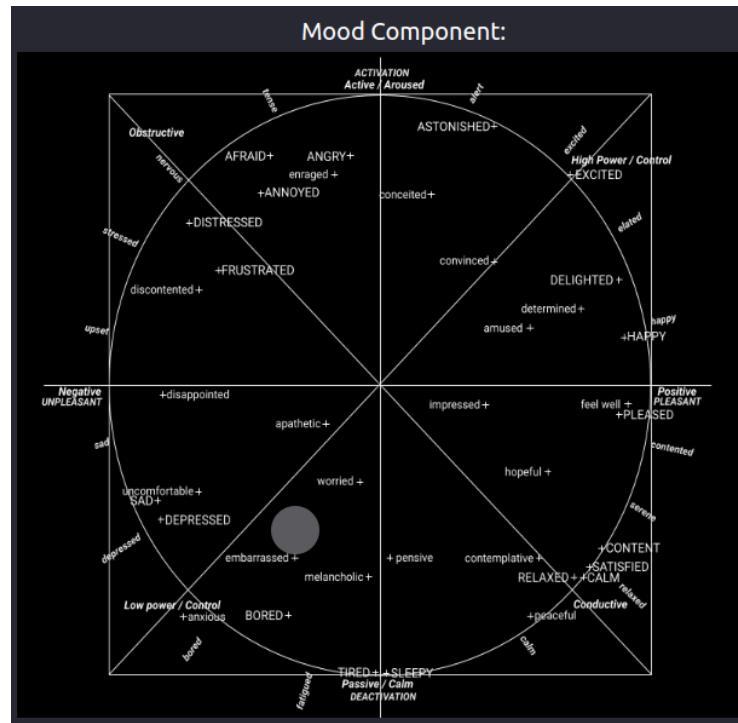


Figura 69: Cuadrante de emociones arrojado por MorphCast para el minuto 0:02 durante la prueba de campo.

Por último una tabla comparativa con los resultados obtenidos en ambas herramientas para el minuto 0:02 del video reacción de la prueba de campo.

	Emoción predominante - Prob.	Valencia	Excitación
Nuestro sistema	Sorpresa - 0.879	-0.118	-0.43
Morph Cast	Sorpresa - 0.77	-0.25	-0.46

Tabla 6: Tabla comparativa para los resultados de ambos sistemas en el minuto 0:02 de la prueba de campo.

Continuando con el minuto 0:12, vemos que los resultados también son considerablemente similares.

Herramienta para la evaluación de emociones en contextos abiertos

En la figura 70 podemos observar que la emoción predominante es miedo, continuando con sorpresa como la siguiente más probable. Este resultado no dista significativamente del resultado de nuestro sistema en la figura 64, el cual detectó sorpresa como emoción predominante, continuando con miedo como la siguiente emoción más probable. Si bien la emoción predominante no es la misma en ambos sistemas, la diferencia es muy sutil, ya que en los dos emociones, las probabilidades rondan el 0.5 en ambas herramientas.

En cuanto a los valores de excitación y valencia, nuestro sistema obtuvo una valencia de -0,484 y una excitación de -0,363, como puede observarse en la figura 64, mientras que el sistema de MorphCast obtuvo una excitación de -0,360, siendo la diferencia despreciable, y una valencia de -0,270. La diferencia respecto de la valencia puede explicarse debido a la diferencia en cuanto a la predominancia de las emociones básicas de un sistema y el otro, como se detallo en el párrafo anterior.

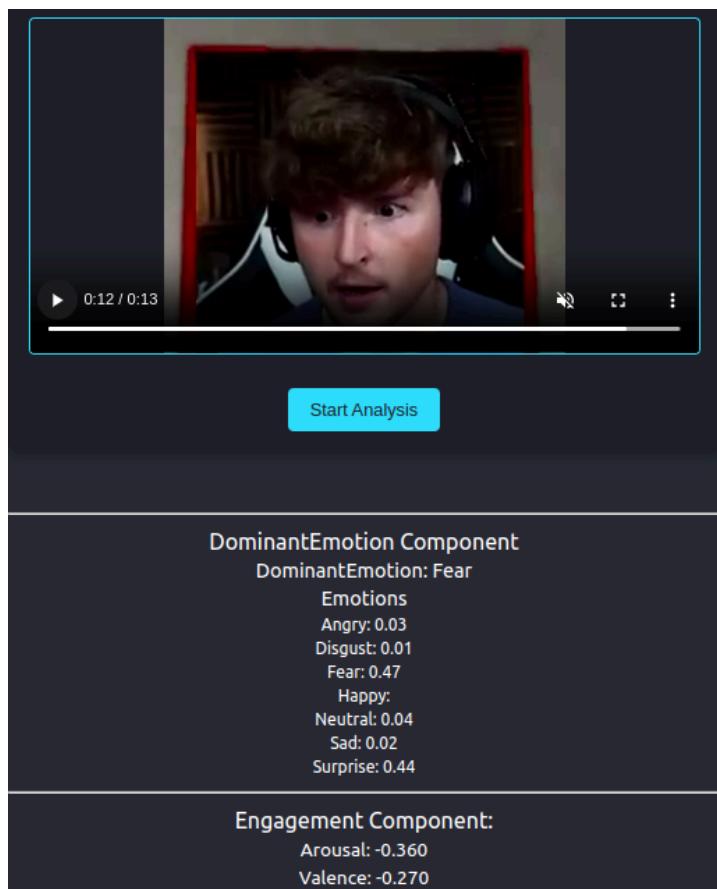


Figura 70: Prueba de campo realizada con MorphCast, muestra del minuto 0:12.

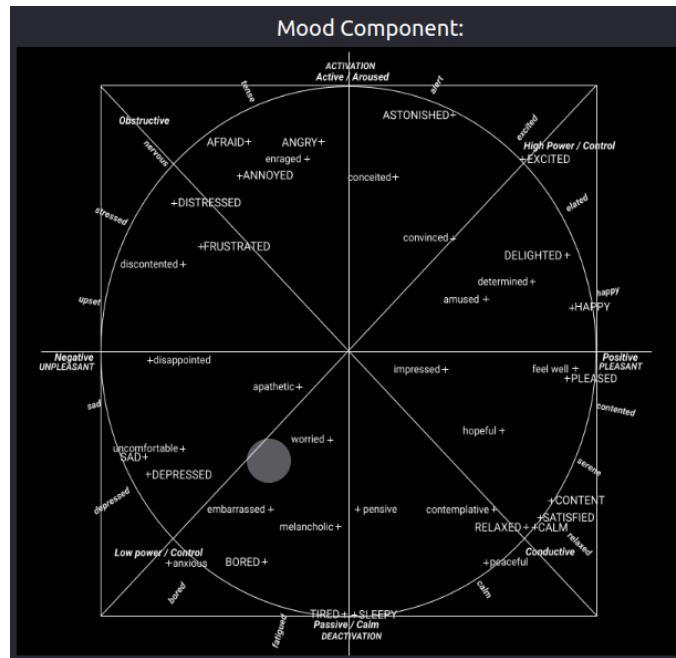


Figura 71: Cuadrante de emociones arrojado por MorphCast para el minuto 0:12 durante la prueba de campo.

Por último una tabla comparativa con los resultados obtenidos en ambas herramientas para el minuto 0:12 del video reacción de la prueba de campo.

	Emoción predominante - Prob.	Valencia	Excitación
Nuestro sistema	Sorpresa - 0.506	-0,484	-0,363
Morph Cast	Miedo - 0,47	-0,270	-0,360

Tabla 7: Tabla comparativa para los resultados de ambos sistemas en el minuto 0:12 de la prueba de campo.

Como detallamos anteriormente, si bien la emoción predominante no es la misma, es algo despreciable ya que la diferencia entre la primera y segunda emoción detectada es mínima en ambos casos, como vimos estas emociones son miedo y sorpresa en ambos sistemas, con casi 0.5 de probabilidad cada una en ambas herramientas.

6.4.2 Análisis de reacciones a videos de DEVO

A continuación, se detallan los resultados de ejecución de MorphCast de algunos de los videos de reacciones de DEVO, previamente analizados en la sección “*6.2 Pruebas de excitación y valencia comparadas con dataset DEVO*” con el fin de comparar resultados con el sistema desarrollado.

6.4.2.1 Video 28.1 - Valentina



Figura 72: Resultados del minuto 0:03 de la herramienta desarrollada para la reacción de Valentina al video 28.1 de DEVO.

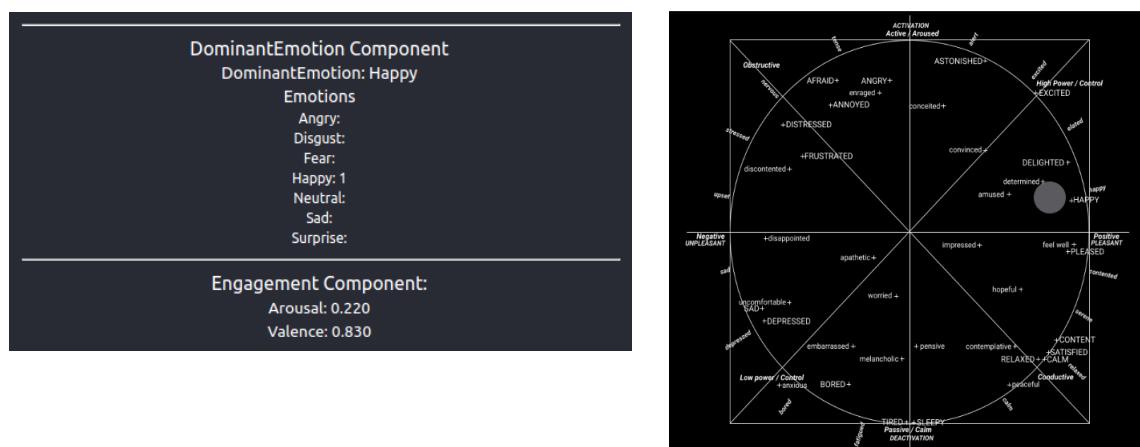


Figura 73: Resultados del minuto 0:03 de la herramienta Morphcast para la reacción de Valentina al video 28.1 de DEVO.

Los resultados obtenidos con la herramienta Morphcast para la reacción de Valentina se aproximan ampliamente. Se puede ver que para ambas corridas, en las figuras 72 y 73, la

Herramienta para la evaluación de emociones en contextos abiertos

emoción básica predominante es felicidad donde la probabilidad obtenida es 1 en ambas herramientas, y la emoción resultante en el circunflejo de Russell se encuentra en el primer cuadrante, ya que en ambos casos obtenemos un valor alto de valencia y un valor intermedio de excitación. Estos valores son de 1 para la valencia y 0.159 para la excitación utilizando nuestro sistema, mientras que MorphCast arroja 0.830 y 0.220 para dichos campos.

6.4.2.2 Video 21.6 - Roma

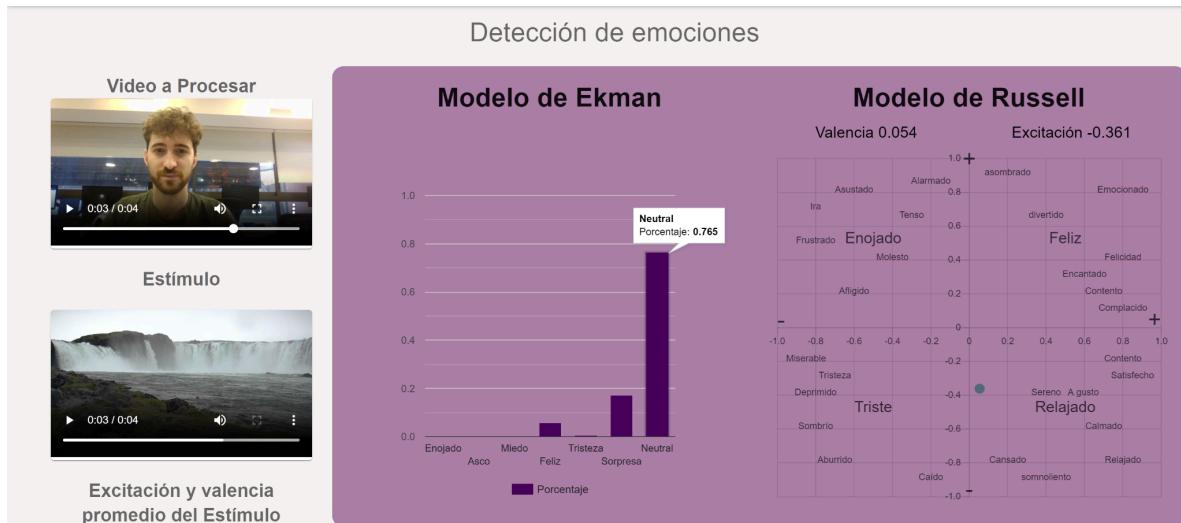


Figura 74: Resultados del minuto 0:03 de la herramienta desarrollada para la reacción de Roma al video 21.6 de DEVO.

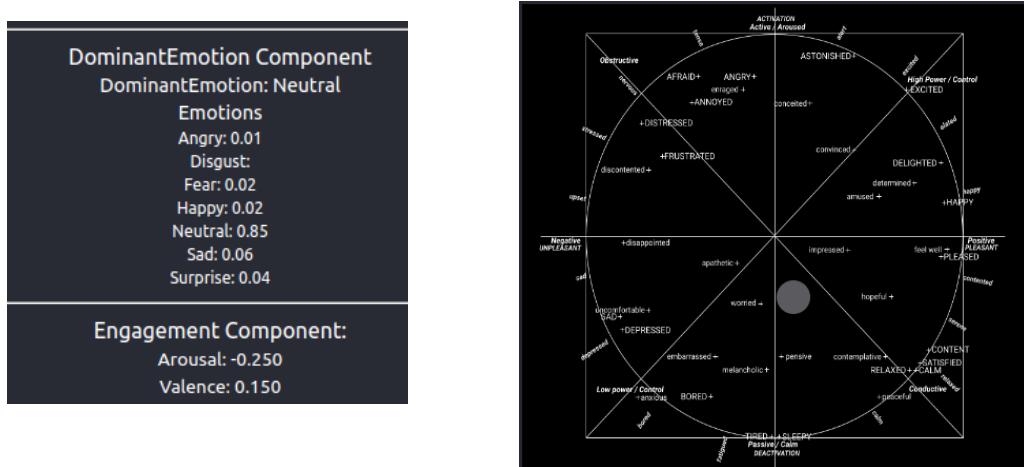


Figura 75: Resultados del minuto 0:03 de la herramienta Morphcast para la reacción de Roma al video 26.1 de DEVO.

En el caso de Roma, los resultados obtenidos para su reacción al video 28.1 también se asemejan considerablemente y lo podemos observar en las figuras 74 y 75. Vemos que la emoción predominante fue neutral, dando una probabilidad de 0.765 para nuestro sistema y 0.85 en MorphCast, y que el cuarto cuadrante predomina en el circunflejo de Russell con valores de excitación y valencia cercanas a cero en ambos sistemas. Estos valores de excitación y valencia fueron de -0.362 y 0.054 para nuestro sistema, y de -0.250 y 0.150 para MorphCast.

6.4.2.3 Video 96.1 - Deborah

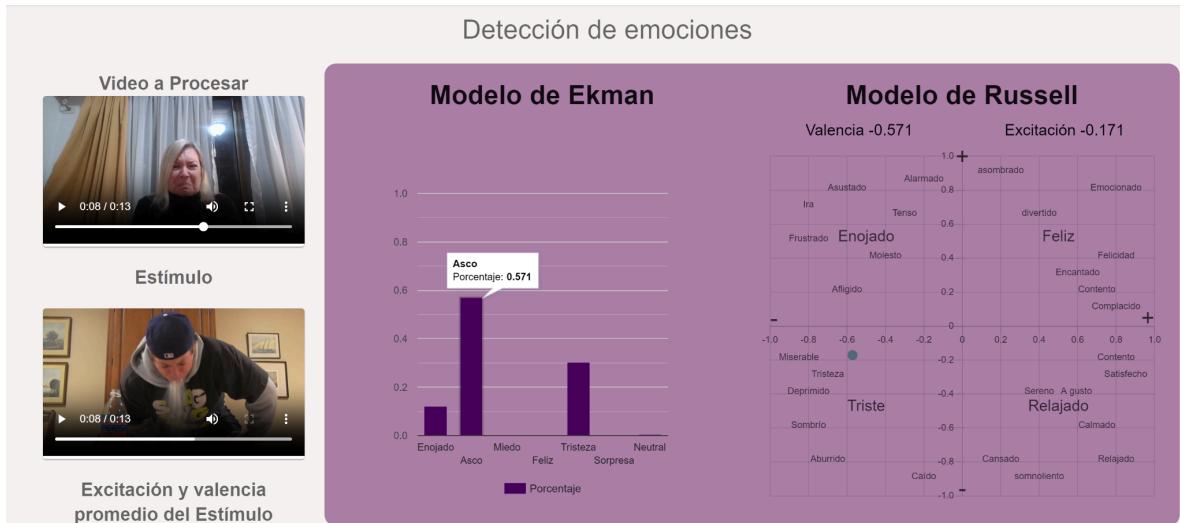


Figura 76: Resultados del minuto 0:08 de la herramienta desarrollada para la reacción de Deborah al video 96.1 de DEVO

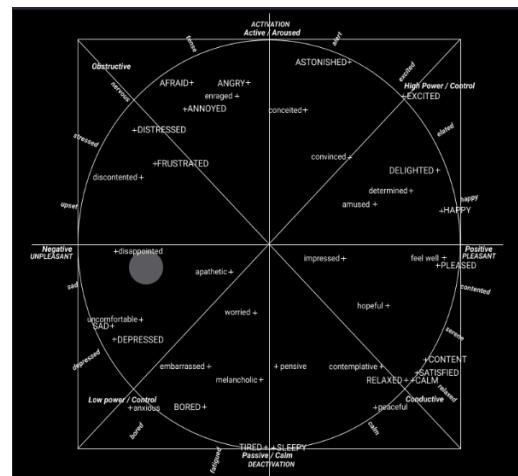
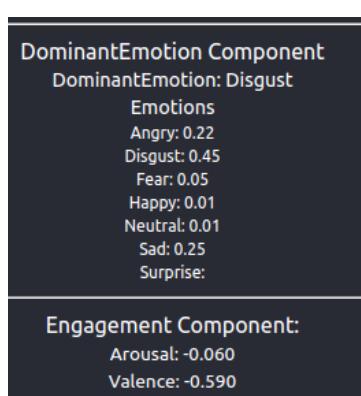


Figura 77: Resultados del minuto 0:08 de la herramienta Morphcast para la reacción de Deborah al video 96.1 de DEVO.

Por último, al analizar con la herramienta de MorphCast la reacción de Deborah al video 96.1 podemos ver en las figuras 76 y 77 que la emoción básica predominante en ambos sistemas fue asco, siendo ligeramente más alta la probabilidad en nuestro sistema, 0.571 contra 0.45. En cuanto al circunflejo de Russell, se obtuvo el tercer cuadrante como el predominante,

con una excitación de -0.175 en nuestro sistema, algo menor que el -0.06 obtenido en MorphCast. En cuanto a la valencia se obtienen valores similares, de -0.571 y -0.590, respectivamente.

6.4.3 Validación con MorphCast - conclusiones

Luego de procesar las reacciones con la herramienta de MorphCast, podemos observar que los resultados son muy similares y comparables con los resultados de las pruebas con el sistema que desarrollamos.

Para el caso de la prueba de campo, nos permitió validar que los resultados de nuestro sistema se encontraban dentro de los valores esperados, particularmente como objetivo primario la validación demostró la pertinencia en la determinación del cuadrante emocional en cada caso presentado.

Para el caso de los videos del set de datos DEVO, y como ya mencionamos previamente, los resultados promedios que provee el mismo pueden ser un tanto diferentes a los obtenidos por nuestro sistema, principalmente por utilizar una forma diferente para calcularlos. MorphCast nos permitió validar y respaldar los resultados del sistema, ya que ante las reacciones a los estímulos DEVO arroja valores muy similares a los nuestros, los cuales también son un tanto diferentes a los promedios de DEVO por lo mencionado anteriormente.

7. Cronograma de las actividades realizadas

En esta sección detallaremos cómo fue el avance del proyecto en el tiempo, especificando los hitos de avance, la matriz de tiempos, y las tareas realizadas mes a mes junto a los integrantes del equipo que realizaron las mismas y el tiempo que consumió cada una de ellas.

7.1 Hitos de avance

- Hito 1: Preparación de los datos. Esto incluye:
 - Investigación de datasets
 - Recolección de los datasets elegidos
 - Preprocesamiento de los datos para el modelo categórico.
- Hito 2: Desarrollo del modelo categórico. Los pasos son:
 - Elaboración
 - Validación
 - Fine tuning
- Hito 3: Desarrollo del modelo dimensional. Para eso se deberá
 - Calcular la valencia utilizando la intensidad de las emociones proporcionadas por el modelo categórico
 - Calcular las unidades de acción por medio de landmarks
 - Calcular el arousal utilizando las unidades de acciones previamente calculadas
- Hito 4: Análisis y validación del modelo dimensional
- Hito 5: Desarrollo del sistema. Los pasos son:
 - Desarrollo de la interfaz de usuario. Entre algunas funcionalidades encontramos:
 - Gráficos de emociones del modelo dimensional
 - Gráficos de valencia y excitación
 - Información sobre el estímulo utilizado
 - Información detallada por individuo y en promedio de personas que aparecen en el video/imagen
 - Desarrollo de la api
 - Integración con el modelo desarrollado
- Hito 6: Validación y despliegue del sistema
- Hito 7: Preparación y publicación de la documentación y artículo científico.

7.2 Matriz de tiempo de los Hitos.

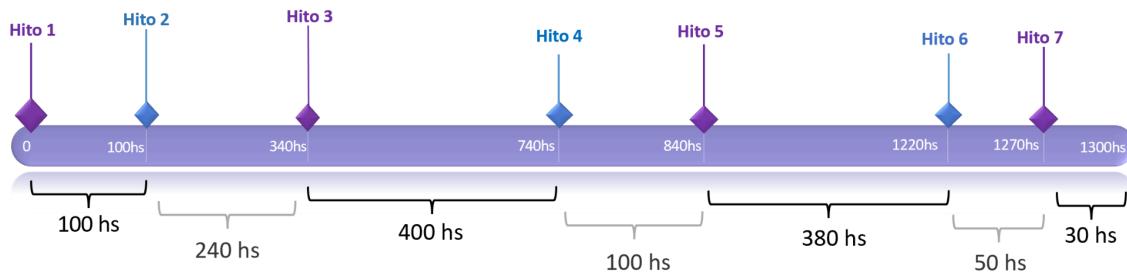


Figura 78: Planificación del tiempo asignado a cada hito.

7.3 Tareas realizadas por mes

Durante el desarrollo de nuestro proyecto profesional, observamos que las tareas se agrupan en ciclos mensuales claramente definidos. En el anexo se encuentra el detalle completo de las tareas realizadas por el equipo, especificando la tarea en sí, la semana en que se llevó a cabo y los integrantes responsables de cada una.

8. Riesgos materializados y lecciones aprendidas

Durante el desarrollo del proyecto, nos enfrentamos a diversos riesgos que impactaron en la ejecución y el progreso del trabajo. Algunos de estos riesgos se materializaron, resultando en desafíos técnicos y de gestión que requirieron adaptaciones y soluciones creativas.

En esta sección, se describen los principales riesgos encontrados y las lecciones aprendidas a lo largo del proceso.

8.1 Riesgos

Uno de los principales riesgos con los que nos encontramos estuvo relacionado con los modelos de machine learning. Dado que nos vimos en la necesidad de utilizar herramientas ya existentes, tuvimos que limitarnos y "adaptarnos" a los modelos desarrollados por terceros. Esto implica, por un lado, que la performance podría no ser óptima en términos de precisión, procesamiento o interfaz de usuario. Además, esta limitación nos restringe a utilizar únicamente las predicciones que proporcionan los modelos. A modo de ejemplo, el modelo de emociones básicas de PyFeat que empleamos no considera la emoción de desprecio, y el modelo de unidades de acción no contempla ciertas unidades, como la 43.

Otra limitación importante fue el uso de diferentes lenguajes de programación. Aunque Python es conveniente para el desarrollo de APIs y machine learning, el primer modelo que utilizamos para la obtención de unidades de acción (OpenGraphAU) tenía una baja performance en cuestiones de tiempo y es por eso que decidimos utilizar (OpenFace) que performaba bien en términos de tiempo y predicción pero estaba implementado en C++. Este no hubiera sido nuestro lenguaje de preferencia debido a su complejidad y al hecho de que no lo habíamos utilizado recientemente. Esto representó un riesgo, ya que el uso de un lenguaje con el que no estábamos familiarizados podría habernos demorado más tiempo debido a la curva de aprendizaje. A su vez la mezcla entre ambos lenguajes para una misma API presentó desafíos a la hora de poder comunicar las distintas partes respetando formatos de tipado y protocolos de comunicación entre ambas.

Otro de los riesgos que se llegó a materializar fue la falta de hardware adecuado para ejecutar todo el sistema. El uso de modelos de Machine Learning, sumado a los distintos servicios como RabbitMQ, Redis y MongoDB provocó que el consumo de recursos, cpu y memoria, sea considerablemente alto logrando así que no cualquier PC pudiera correr el sistema en condiciones.

En relación al punto anterior, también encontramos dificultades a la hora de querer desplegar nuestro sistema en entornos cloud como habíamos propuesto originalmente. No por limitaciones en cuanto a conocimientos, sino por los altos costos de instanciar estos servicios. Gracias a créditos otorgados por ciertas plataformas pudimos realizar pruebas de despliegue en

Kubernetes, pero estos créditos no fueron suficientes como para permitirnos dejar el sistema desplegado y funcionando productivamente.

8.2 Lecciones Aprendidas

A lo largo del desarrollo del proyecto, aprendimos varias lecciones valiosas. En primer lugar, la importancia de la flexibilidad y adaptabilidad en el uso de herramientas y tecnologías. Aunque inicialmente estábamos limitados por los modelos existentes y los lenguajes de programación seleccionados, estas restricciones nos llevaron a explorar nuevas formas de optimización y mejora, lo cual enriqueció nuestro conocimiento y habilidades.

En segundo lugar, la planificación anticipada y la evaluación de riesgos técnicos son cruciales. La necesidad de hardware adecuado y la elección de lenguajes de programación no solo afectan el desarrollo, sino también la viabilidad y eficiencia del proyecto. En el futuro, será fundamental realizar una evaluación más detallada de los recursos disponibles y las tecnologías a utilizar desde las primeras fases del proyecto.

Otra gran lección aprendida fue el hecho de desarrollar un sistema que debe procesar una gran cantidad de información lo más rápido posible. Durante el desarrollo del proyecto fue necesario adaptarnos a manipular imágenes y videos, lo cual es muy diferente a trabajar con otros tipos de datos o archivos como lo pueden ser el texto y los audios. Esto nos llevó a explorar diversas herramientas y tecnologías, tales como OpenCV, Pillow (PIL), y modelos de machine learning como PyFeat, FerPlus o OpenFace con el fin de poder aplicar mejores prácticas y así optimizar el proceso.

Finalmente, comprendimos la relevancia de una gestión eficaz del tiempo y de la curva de aprendizaje asociada a nuevas tecnologías. Aunque fue desafiante, esta experiencia nos enseñó a abordar nuevos lenguajes y herramientas de manera más estratégica, aprovechando la documentación y los recursos disponibles para acelerar el proceso de aprendizaje y adaptación.

9. Trabajos futuros

En esta sección se describen posibles mejoras y extensiones que podrían ser implementadas en futuras versiones de nuestro sistema de detección de emociones para imágenes y videos utilizando el modelo categórico y dimensional. Estas propuestas buscan ampliar las funcionalidades y la precisión del sistema, así como su aplicabilidad en diversos entornos.

En primer lugar mencionaremos puntos vinculados a la implementación y luego posibles adiciones relacionadas al dominio del problema.

Una mejora significativa sería la incorporación de soporte para la grabación de imágenes o videos en tiempo real, permitiendo que el sistema no solo detecte emociones de videos pregrabados, sino también a medida que una persona se graba en vivo. Esto proporcionaría una herramienta más dinámica y adaptable para aplicaciones como entrevistas, monitoreo en tiempo real en entornos educativos o de salud, y experiencias interactivas. Esto está muy relacionado con el hardware con el que se cuenta para correr el sistema, probar el uso de GPUs para acelerar las predicciones podría ser muy beneficioso.

Relacionado al punto anterior se encuentra la robustez de la arquitectura. En sistemas distribuidos como el nuestro, es muy importante implementar una lógica para que los componentes sean tolerantes a fallos y brinden alta disponibilidad. Esto también está sumamente relacionado con la implementación de una lógica que permita aumentar o disminuir las réplicas de forma automática, utilizando métricas como pueden ser cantidad de usuarios conectados, carga en nuestras colas, etc.

Continuar evaluando y probando otros modelos de machine learning para detección de emociones, unidades de acción e incluso modelos de cálculo de valencia y excitación. Explorar nuevas arquitecturas de redes neuronales, técnicas de aprendizaje profundo y conjuntos de datos actualizados para mejorar la precisión y la robustez del sistema en diversas condiciones y con diferentes poblaciones podría brindar una mejora significativa en el sistema, haciéndolo más personalizable y, probablemente, más eficaz.

Otro punto de mejora consistiría en entrenar modelos de Machine Learning propios y lograr estandarizar los dos procesadores en un solo, capaz de hacer el análisis completo de cada frame retornando la valencia y excitación de los mismos.

En cuanto posibles agregados relacionados a la “Computación afectiva”, se podría incorporar la metodología de encuestas SAM para la obtención de valencia y excitación de un estímulo. La combinación de datos objetivos de detección facial con respuestas subjetivas de los usuarios proporciona una evaluación más holística y precisa de las emociones.

A su vez se podría desarrollar funcionalidades que permitan el seguimiento de los máximos y mínimos históricos de las emociones detectadas a lo largo del video. Esto permitiría un análisis temporal detallado, identificando momentos de mayor impacto emocional y

Herramienta para la evaluación de emociones en contextos abiertos

proporcionando insights más profundos sobre la dinámica emocional en el transcurso del video.

Finalmente creemos que un análisis multimodal combinando datos faciales con otros indicadores emocionales, como el tono de voz y el lenguaje corporal, permitiría una evaluación más completa y precisa de las emociones, y la posibilidad de utilizar el sistema en nuevo contextos, especialmente en aquellos donde la expresión facial por sí sola no es suficiente.

10. Conclusiones

Para este trabajo implementamos un sistema capaz de reconocer emociones a partir de una imagen o video dentro de un contexto dado, empleando una combinación entre el modelo categórico y el modelo dimensional. Esto surge a partir de que hoy en día no hay sistemas gratuitos o de código libre que proporcionen un análisis en profundidad sobre la detección de emociones utilizando ambos modelos en conjunto.

Durante el desarrollo del proyecto nos vimos en la necesidad de aplicar muchos de los conocimientos y habilidades adquiridos a lo largo de la carrera. Desde el desarrollo básico de software, problemas de concurrencia y sistemas distribuidos hasta aspectos de gestión y planeamiento de proyectos, pasando también por el análisis de datos y todo lo que ello conlleva. Creemos que la carrera contribuyó fuertemente en nuestra habilidad para adaptarnos a trabajar con nuevas herramientas y frameworks que no conocíamos, buscar soluciones creativas a los problemas y riesgos mencionados anteriormente y, principalmente, a ser capaces de determinar, diseñar y priorizar las tareas a desarrollar, junto con el impacto que estas tendrían, previo al desarrollo de las mismas. Esta habilidad nos permitió entender a qué situaciones y/o fases del proyecto debemos destinar un mayor esfuerzo y tiempo y a cuáles no.

Por otro lado, entendimos como la Ingeniería en Informática puede ser aplicada a diversos problemas con dominios totalmente diferentes unos de otros. En nuestro caso este dominio fue la ‘Computación afectiva’, área en la cual no contábamos con experiencia previo al desarrollo del trabajo y es por eso que fue muy importante dedicar los primeros meses a investigación y entendimiento del mismo, creemos que la carrera también contribuye en este aspecto y nos brinda la capacidad de entender nuevos conceptos y problemas rápidamente.

Todo lo mencionado anteriormente demuestra cómo el Trabajo Profesional afecta y aporta a nuestra formación dado que permite aplicar en forma conjunta muchos de los conocimientos y habilidades que durante la carrera adquirimos de forma “aislada” o en diferentes materias.

Como describimos en la Sección 6 acerca de los resultados obtenidos, fuimos capaces de validar nuestro sistema realizando comparaciones con otros frameworks similares, utilizando videos e imágenes de datasets destinados a ello, y también realizando pruebas de campo. Los resultados obtenidos fueron satisfactorios, llegamos a la conclusión de que es posible aplicar ambos modelos, tanto el categórico como el dimensional, de forma de obtener un análisis más preciso de las emociones en las personas.

Nuestro sistema está diseñado para reconocer emociones en ‘*contextos abiertos*’, es decir que puede ser utilizado en diversos ámbitos. Como por ejemplo, la reacción de alumnos ante un clase universitaria, el entrenamiento de profesionales en uso de simuladores, el diseño de videojuegos y como las diferentes escenas afectan a los usuarios, el análisis en reuniones virtuales, entre muchos otros. Es por eso que, al no existir otras herramientas de uso libre que

Herramienta para la evaluación de emociones en contextos abiertos

realicen este análisis en conjunto, creemos que nuestra implementación puede ser de mucha utilidad para la comunidad, tanto técnica como no técnica.

Si bien los resultados fueron satisfactorios, como ya mencionamos en la Sección 9 creemos que existen puntos de mejora y margen para continuar con la investigación y desarrollo del sistema. Especialmente teniendo en cuenta la gran adopción e importancia que la Inteligencia artificial y los modelos de Machine Learning están tomando con el paso del tiempo, acompañado con el uso cada vez mayor de dispositivos electrónicos e interacciones interpersonales de forma virtual.

Referencias

- [1] Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, 44(4), 695–729. <https://doi.org/10.1177/0539018405058216>
- [2] *Universal Emotions | What are Emotions? | Paul Ekman Group*. (2023, May 1). Paul Ekman Group. <https://www.paulekman.com/universal-emotions/>
- [3] Ekman, P., Friesen, W. V., O’Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., Scherer, K., Tomita, M., & Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712–717. <https://doi.org/10.1037/0022-3514.53.4.712>
- [4] Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200. <https://doi.org/10.1080/0269939208411068>
- [5] Baldasarri S. *Computación Afetiva: tecnología y emociones para mejorar la experiencia de usuario*. Recuperado online
[http://sedici.unlp.edu.ar/bitstream/handle/10915/53441/Documento_completo_.pdf-PDFA.pdf?sequence=1&isAllowed=y]
- [6] Fellous, J. M. (2004). From human emotions to robot emotions. *National Conference on Artificial Intelligence*. <http://amygdala.psychdept.arizona.edu/pubs/AAAI2004-reprint.pdf>
- [7] Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129. <https://doi.org/10.1037/h0030377>
- [8] Ekman, P., & Friesen, W. V. (1978, January 1). *Facial Action Coding System*. PsycTESTS Dataset. <https://doi.org/10.1037/t27734-000>
- [9] Barrionuevo, C., Ierache, J. S., & Sattolo, I. I. (2020). *Reconocimiento de emociones a través de expresiones faciales con el empleo de aprendizaje supervisado aplicando regresión logística*. <http://sedici.unlp.edu.ar/handle/10915/114089>
- [10] Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- [11] Jorge, I., Iris, S., & Gabriela, C. (2020). Framework multimodal emocional en el contexto de ambientes dinámicos. *RISTI*, 40, 45–59. <https://doi.org/10.17013/risti.40.45-59>
- [12] Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), 102–141. <https://doi.org/10.1033-2909.115.1.102>
- [13] *Detección de rostros*. (2024, marzo 8). Google for Developers. Último Acceso junio 19, 2024, de <https://developers.google.com/ml-kit/vision/face-detection?hl=es-419>
- [14] Face++ (2024) <https://console.faceplusplus.com>

- [15] Onnx (2024). Emotion Ferplus. Models.
https://github.com/onnx/models/tree/main/vision/body_analysis/emotion_ferplus
- [16] GitHub - Microsoft/FERPlus: This is the FER+ new label annotations for the Emotion FER dataset. (2017, abril 25). GitHub. Último acceso junio 19, 2024,
<https://github.com/microsoft/FERPlus>
- [17] ONNX | Home. (2019). <https://onnx.ai/>
- [18] Amazon Rekognition. (2023). Amazon Web Services, Inc.; Amazon.
<https://aws.amazon.com/es/rekognition/>
- [19] Detecting faces in an image - Amazon Rekognition. (2024). Retrieved June 19, 2024, from
<https://docs.aws.amazon.com/rekognition/latest/dg/faces-detect-images.html>
- [20] Abuhammad, H., & Everson, R. (2018). Emotional Faces in the Wild: Feature Descriptors for Emotion Classification. In *Lecture notes in computer science* (pp. 164–174).
https://doi.org/10.1007/978-3-319-93000-8_19
- [21] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression.* <https://doi.org/10.1109/cvprw.2010.5543262>
- [22] Loijens, L., & Krips, O. (2021). *FaceReader Methodology Note*. Retrieved June 19, 2024, from
https://info.noldus.com/hubfs/resources/noldus-white-paper-facereader-methodology.pdf?utm_campaign
- [23] Mollinedo, D. L. (2019, junio 26). *Api rest para el reconocimiento facial de emociones (Fer Rest Api)*. <https://dspace.uclv.edu.cu/handle/123456789/12159>
- [24] Gavrilov V., “Registro Emocional”, Universidad de Buenos Aires, FIUBA
█ Registro emocional -Herramientas Modelo Categorico.pdf
- [25] Labeled Faces in the Wild (LFW) Dataset. (2018, mayo 17). Kaggle.
<https://www.kaggle.com/datasets/jessicali9530/lfw-dataset>
- [26] MorphCast | World's Smallest & Greenest Facial Emotion Recognition AI. (2024, Junio 3). MorphCast. <https://www.morphcast.com/>
- [27] Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- [28] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression.* <https://doi.org/10.1109/cvprw.2010.5543262>
- [29] What are the advantages and disadvantages of the different emotion theories and models? | 5 Answers from Research papers. (n.d.). SciSpace - Question.

- <https://typeset.io/questions/what-are-the-advantages-and-disadvantages-of-the-different-47le0xyqkf>
- [30] Barrett, L. F. (2006). Are Emotions Natural Kinds? *Perspectives on Psychological Science*, 1(1), 28–58. <https://doi.org/10.1111/j.1745-6916.2006.00003.x>
- [31] Lang, P. J. (1995). *International Affective Picture System (IAPS) : Technical Manual and Affective Ratings*. <https://ci.nii.ac.jp/naid/10010070032>
- [32] Miccoli, L. (2016, June 1). *OLAF, the Open Library of Affective Foods in ADULTS*. Datasets. <https://doi.org/10.30827/digibug.41499>
- [33] Ack Baraly, K. T., et al. (2020). *Database of Emotional Videos from Ottawa (DEVO)*. Collabra: Psychology, 6(1): 10. DOI: <https://doi.org/10.1525/collabra.180>
- [34] Alonso M., González S. *Reconocimiento de emociones a través de sensores fisiológicos con el empleo de aprendizaje supervisado reforzado por emociones faciales*.
- [35] Asyncio — Asynchronous I/O. (2015, marzo 10). Python Documentation. ultimo accesso junio 19, 2024, de <https://docs.python.org/es/3/library/asyncio.html>
- [36] Mosquito. (2024, mayo 21). *GitHub - mosquito/aiormq: Pure python AMQP 0.9.1 asynchronous client library*. GitHub. <https://github.com/mosquito/aiormq>
- [37] Motor: Asynchronous Python driver for MongoDB — Motor 3.4.0 documentation. (2024). <https://motor.readthedocs.io/en/stable/>
- [38] (Asyncio OR Threadsafe) Google Cloud Client Libraries for Python. (2024). <https://talkiq.github.io/gcloud-aio/>
- [39] TadasBaltrusaitis. (2018). *GitHub - TadasBaltrusaitis/OpenFace: OpenFace – a state-of-the art tool intended for facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation*. GitHub. <https://github.com/TadasBaltrusaitis/OpenFace>
- [40] JWT.IO. (n.d.). JSON Web Tokens - jwt.io. <https://jwt.io/>
- [41] Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- [42] Ack Baraly, K. T., Allard, L. A., Cross, E. C., Goupil, L., & Deslandes, M. (2020). Database of Emotional Videos from Ottawa (DEVO). *Collabra: Psychology*, 6(1), Article 10. <https://doi.org/10.1525/collabra.180>
- [43] Lang, P. (2005). International affective picture system (IAPS) : affective ratings of pictures and instruction manual. *CTIT Technical Reports Series*. <https://ci.nii.ac.jp/naid/20001061266>
- [44] RabbitMQ: One broker to queue them all | RabbitMQ. (2024). <https://www.rabbitmq.com/>
- [45] FastAPI. (2024). <https://fastapi.tiangolo.com/>

- [46] Lingjivoo. (2023, Mayo 30). GitHub - lingjivoo/OpenGraphAU: A tool for facial action unit analysis. GitHub. <https://github.com/lingjivoo/OpenGraphAU>
- [47] 3.1 Coeficiente de Pearson. (n.d.). Último acceso junio 19, 2024. https://www.uv.es/webgid/Descriptiva/31_coeficiente_de_pearson.html
- [48] Fowler, M. (2020, Marzo 30). blik: Pair Programming. martinfowler.com. <https://martinfowler.com/bliki/PairProgramming.html>
- [49] What is a REST API? | IBM. (n.d.). Último acceso junio 19, 2024. <https://www.ibm.com/topics/rest-apis>
- [50] Caylus. (2023, Abril 10). I Played a BANNED Horror Game.. [Video]. YouTube. <https://www.youtube.com/watch?v=pNDg3yyZP10>
- [51] Farnsworth, B. (2024, mayo 24). Facial Action Coding System (FACS) - A Visual Guidebook - iMotions. iMotions. <https://imotions.com/blog/learning/research-fundamentals/facial-action-coding-system/#emotions-and-action-units>
- [52] Cosenlab. 2024. Py-feat: Python facial expression analysis toolbox. <https://py-feat.org/pages/intro.html>

Anexos

Manual de ejecución

Para ejecutar el sistema se puede consultar la guía en:

<https://trabajo-profesional-grupo-21.github.io/manual-ejecucion/>

Código fuente del sistema

El código fuente del sistema se encuentra en la sección ‘Repositorios’ de la siguiente organización de GitHub: <https://github.com/Trabajo-profesional-grupo-21>

Comparación de frames a procesar

El desarrollo de las pruebas para realizar la comparación de la cantidad de frames procesados se encuentra en el siguiente notebook y carpeta  Comparativa_emociones.ipynb

 1 vs all - Frames

Comparativa FerPlus - PyFeat

Hoja de cálculos con la comparativa en las predicciones de emociones para los modelos de FerPlus y PyFeat  Comparativa FerPlus - PyFeat

Comparativa unidades de acción

Hoja de cálculos con la comparativa en las predicciones de unidades de acción para los modelos de OpenFace, OpenGraphAU y Noldus  Validacion AU

Pruebas de excitación y valencia

Los videos y los análisis realizados durante las pruebas de excitación y valencia se encuentran en el siguiente link  videos .

En la carpeta de “Análisis realizados” se encuentran los resultados de los análisis realizados con el sistema mientras que en la carpeta de “Videos estímulo devo” se encuentran los videos del dataset de devo seleccionados para realizar dichas pruebas.

Pruebas de Campo

El video utilizado para procesar junto a su estímulo y su análisis se encuentra en el siguiente link  Prueba de campo

Tareas realizadas por mes

A continuación se presenta un informe formal sobre las tareas realizadas por el equipo durante todo el desarrollo del proyecto. Las tareas se encuentran agrupadas por mes en tablas, donde se detalla la tarea en sí, la semana en que se realizó dicha tarea y los integrantes que la llevaron a cabo. Para acceder al detalle completo de las tareas, puede consultar el siguiente enlace: [+ Detalle de Tareas](#).