



**BENEMERITA UNIVERSIDAD AUTONOMA DE
PUEBLA**

Facultad de Ciencias de la Computación

MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

Tesis:

*“ALGORITMO PARA DETERMINAR EL NIVEL DE
ATENCIÓN DE UN USUARIO EN UN SISTEMA DE
CÓMPUTO BASADO EN VISIÓN COMPUTACIONAL”*

Presenta:

Nallely Morales Lozada

Directores:

José Luis Carballido Carranza

Iván Olmos Pineda

Abril del 2022.

Dedicatoria.

Primeramente a Dios, por darme la fortaleza durante tantos años para no darme por vencida frente a ningún obstáculo.

A mi mamá por su apoyo incondicional tanto en mi crecimiento personal como profesional, por creer en mí e impulsarme a lograr lo inimaginable, por estar en cada etapa de mi vida, celebrar y alegrarse de mis logros como si fuesen suyos. Gracias por inculcar en mí el ejemplo de esfuerzo y la perseverancia.

A mi padre, a quien con su amor, paciencia y esfuerzo me ha permitido cumplir una meta más, por dar todo de sí mismo para nosotros, por esforzarse en cada paso para demostrarnos lo grandioso que es, por creer siempre en mí y en mis sueños, por amarme pese a todo.

A mi hijo por motivarme a salir adelante, por demostrarme que todo es posible con un poco de ayuda de quienes amas, por creer y crecer conmigo durante mi formación académica. Por darme las palabras más simples y llenas de amor que lograron que cualquier momento de agotamiento quedara atrás.

A mi prometido, por estar conmigo en mis momentos de debilidad, por darme las palabras más reconfortantes y sobre todo por saber escucharme y hacer que cualquier momento malo pasara rápido. Por su gran amor, total apoyo y comprensión a lo largo de este tiempo.

A mi hermano, por cuidar de mí y de mi hijo siempre, por estar en los momentos de alegría y en los que no lo son tanto. Por demostrarnos que a pesar de todo siempre estará para nosotros.

A mi ojitos, a quien lejos de yo salvar, me salvo a mí.

Nallely Morales Lozada

Agradecimientos.

Agradezco la oportunidad al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo otorgado para la realización de mis estudios de Maestría.

A la Benemérita Universidad Autónoma de Puebla y al posgrado de la Facultad de Ciencias de la Computación, por recibirme y hacerme participe de su comunidad estudiantil.

Agradezco a mi director de tesis, al Dr. Iván Olmos Pineda por darme la oportunidad de trabajar con él en este proyecto, por siempre resolver mis inquietudes y apoyarme durante el tiempo de la realización de esta tesis.

A mis compañeros de la maestría por su ayuda durante mi estancia en Puebla, a Daniel, Irving, Alberto, Yael y Paco quienes estuvieron apoyándome en todo momento, tanto en los aspectos académicos como personales.

A mi familia y amigos quienes me ayudaron a concluir mi proyecto de tesis, realizando las pruebas necesarias para verificar la eficacia de mi sistema. Gracias a quienes fueron partícipes de mi formación como profesionista.

Resumen.

En el presente proyecto, se muestra una serie de algoritmos empleados para determinar si una persona está prestando atención a lo que observa en una pantalla durante actividades que se llevan a cabo en tiempo real en sistemas de cómputo. En este trabajo, el término “prestar atención” estará relacionado a la actividad en la cual, una persona mantiene en un alto porcentaje de tiempo, su cara en orientación a la posición donde se encuentra posicionada la pantalla. Para ello, se asume que existe una cámara de video ubicada sobre o debajo del monitor, la cual estará orientada a la posición de la persona.

El proyecto tiene como objetivo el ayudar a la toma de decisiones, según sea el caso, en tareas de educación, seguimiento y/o coordinación de tareas que involucren la transmisión de información bidireccional por medio de eventos de video, entre otros, los cuales han tomado una relevancia muy alta derivada de los eventos provocados por la pandemia del COVID-19, en donde el trabajo remoto en línea ha emergido como una actividad primordial en la sociedad moderna.

Contenido.

Dedicatoria.....	I
Agradecimientos.....	II
Resumen.....	III
Capítulo I. Introducción.....	11
1.1 Motivación.....	12
1.2 Objetivos.....	13
1.2.1 Objetivo general.....	13
1.2.2 Objetivos específicos.....	14
1.3 Alcances y limitaciones.....	14
1.3.1 Alcances.....	14
1.3.2 Limitaciones.....	15
1.4 Estructura del documento.....	16
Capítulo II. Estado del arte.....	18
2.1 Detección de rostros.....	18
2.2 Detección de ojos.....	20
2.3 Detección de pose de la cabeza.....	21
2.4 Detección del nivel de atención.....	22
Capítulo III. Marco teórico.....	24
3.1 Reconocimiento facial.....	24
3.2 Algoritmos de detección facial.....	25
3.2.1 Algoritmo de Viola-Jones.....	25
3.2.2 Detección facial basada en redes neuronales.....	26
3.2.3 Puntos de referencia faciales.....	27
3.2.3.1 Dlib.....	27
3.2.3.2 Flandmark.....	28
3.2.4 Extracción de características.....	28
3.2.4.1 Métodos basados en la apariencia (holísticos).....	28
3.2.4.2 Métodos basados en modelos.....	29
3.3 Métodos no invasivos para detección de ojos.....	30
3.3.1 Transformada de Hough circular.....	30

3.3.2 Haar Cascade	30
3.3.3 Dlib.....	30
3.4 Detección de pose	31
3.4.1 Plantillas de apariencia	32
3.4.2 Conjunto de detectores	32
3.4.3 Modelos flexibles.....	32
3.4.3.1 Active Appearance Model (AAM).....	32
3.4.3.2 Active Shape Model (ASM).....	33
3.4.4 Proyección 3D.....	33
3.5 Comportamiento frente al ordenador	33
3.5.1 Distancia entre la pantalla/usuario	34
3.5.2 Postura de la cabeza	34
3.5.3 Iluminación.....	34
3.5.4 Parpadeo	35
3.6 ¿Qué es la atención?.....	35
3.6.1 Factores determinantes de la atención	36
Capitulo IV. Propuesta de solución.....	38
4.1 Desarrollo de la propuesta.....	38
4.2 Diagrama de flujo método propuesto.....	40
4.3 Procesamiento frames	41
4.3.1 Detección facial.....	41
4.3.2 Detección de la relación de aspecto del ojo EAR	41
4.3.3 Calculo del parpadeo	42
4.3.4 Detección de pose	43
4.4 Procesamiento de los resultados.....	44
4.4.1 Relación de aspecto ocular (EAR)	45
4.4.2 Parpadeo	46
4.4.3 Movimientos de la cabeza.....	46
4.5 Determinar nivel de atención	47
4.5.1 Nivel de atención por pose de la cabeza	47
4.5.1.1 Postura final cabeza.....	48
4.5.2 Nivel de atención por EAR.....	49
4.5.3 Nivel de atención por parpadeo	50
4.6 Desarrollo de fase experimental	50

4.6.1 Software.....	50
4.6.2 Hardware	51
4.6.3 Librerías de Python.....	51
Capitulo V. Resultados	53
5.1 Reporte de usuarios/ nivel de atención.....	54
5.1.1 Tarea 1. Clase en línea.....	54
5.1.2 Tarea 2. Lectura de documento academico.....	56
5.1.3 Tarea 3. Actividad de interes	58
5.1.4 Tarea 4. Escritura de documento	60
5.2 Resultados obtenidos	62
5.2.1 Tarea 1.....	62
5.2.2 Tarea 2.....	66
5.2.3 Tarea 3.....	69
5.2.4 Tarea 4.....	73
5.2.5 Análisis de los resultados.....	76
Conclusiones	77
Trabajo futuro	78
Referencias.....	79

Índice de figuras

Figura 1. Representación genérica de una neurona artificial	26
Figura 2. Puntos de referencia faciales presentados por Dlib [20]	28
Figura 3. Ángulos de rotación pitch, roll y yaw [26]	31
Figura 4. Postura correcta para trabajar frente a la computadora [31]	34
Figura 5. Desarrollo de la propuesta.....	38
Figura 6. Puntos de referencia ojo izquierdo	42
Figura 7. Puntos faciales para detección de pose.....	44
Figura 8. Usuario mirando fuera de la pantalla. Yaw= 38°	47
Figura 9. Usuario mirando por encima de la pantalla. Pitch= -9.43°	48
Figura 10. Usuario mirando a la pantalla con flexión lateral. Roll=10.24°	48
Figura 11. Apertura ocular máxima del usuario. EAR=0.36.....	49
Figura 12. Apertura ocular mínima del usuario, sin parpadeo. EAR=0.24	49
Figura 13. Apertura ocular antes, durante y después de un parpadeo	50
Figura 14. Usuario 9 en tarea 1	54
Figura 15. Resultados de usuario 1	56
Figura 16. Resultados del usuario 19.....	58
Figura 17. Parpadeos totales realizados por el usuario	58
Figura 18. Resultados obtenidos en actividad de escritura	60

Índice de tablas.

Tabla 1. Ejemplo de los resultados obtenidos a la finalización del sistema.	45
Tabla 2. Software empleado	51
Tabla 3. Hardware utilizado	51
Tabla 4. Librerías empleadas en el proyecto	52
Tabla 5. Resultados de tarea 1	63
Tabla 6. Parpadeos realizados por los usuarios durante la prueba	66
Tabla 7. Resultados mostrados por el sistema para tarea 2	66
Tabla 8. Nivel de atención de los usuarios en la tarea 3.....	70
Tabla 9. Numero de parpadeos realizados en la 3ra tarea	72
Tabla 10. Resultados obtenidos del sistema para la tarea 4.....	73
Tabla 11. Numero de parpadeos realizados por los usuarios en la 4ta tarea	75

Índice de Graficas.

Gráfica 1. Porcentajes de la cabeza obtenidos por el usuario 19.....	55
Gráfica 2. EAR del usuario 9, durante la tarea 3	55
Gráfica 3. Movimientos de la cabeza del usuario 1	57
Gráfica 4. Variaciones EAR usuario en tarea 2.	57
Gráfica 5. Movimientos de la cabeza realizados por el usuario	59
Gráfica 6. EAR presentado por el usuario	60
Gráfica 7. Movimientos de la cabeza usuario 17	61
Gráfica 8. Relación de aspecto ocular del usuario 17	62
Gráfica 9. Porcentaje de atención de los 20 usuarios	63
Gráfica 10. Tiempo de atención durante la prueba	64
Gráfica 11. Porcentaje de movimientos de la cabeza de la tarea 1	65
Gráfica 12. Porcentaje de EAR de los usuarios.....	65
Gráfica 13. Nivel de atención de los usuarios a la prueba 2.....	67
Gráfica 14. Tiempo de atención de los usuarios.....	67
Gráfica 15. Movimientos de la cabeza de los usuarios	68
Gráfica 16. EAR de los 20 usuarios para la tarea 2	68
Gráfica 17. Nivel de atención 3ra tarea	70
Gráfica 18. Tiempo de atención de los participantes en tarea 3.....	71
Gráfica 19. Porcentaje de movimientos de la cabeza.....	71
Gráfica 20. Porcentaje EAR prueba 3	72
Gráfica 21. Nivel de atención de la tarea 4.....	73
Gráfica 22. Tiempo de atención de los participantes para la 4ta prueba	74
Gráfica 23. Porcentaje de movimiento de la cabeza.....	74
Gráfica 24. Porcentaje de EAR de la 4ta tarea	75

Capítulo I. Introducción

El área de visión computacional, específicamente el procesamiento digital de imágenes, es el conjunto de técnicas computacionales que permiten la extracción de características o patrones para un determinado propósito. Dentro de esta área existen diversas disciplinas como la reconstrucción de escenas para obtener modelos 3D, la restauración de imágenes, el reconocimiento visual, es decir, determinar si un determinado objeto, clase o característica está presente en una imagen, entre otros.

Se han desarrollado diversos algoritmos que le permiten a una computadora capturar la información de su entorno para lograr reconocer o identificar características de los objetos, personas o animales que permiten lograr la detección de patrones, clasificación, procesamiento de imágenes, identificación de áreas de interés, entre otros.

La detección de rostros, es un subdominio de la detección de objetos, uno de los campos basados en la interacción computadora-humano. La detección de objetos es un proceso que se ocupa de detectar instancias en objetos de una clase en particular (como personas, automóviles, edificios o rostros) en una imagen estática o una secuencia de ellas.

Para dichas tareas, existen técnicas específicas que ayudan a procesar e interpretar la información obtenida y determinar en una determinada escena si existe el rostro de una persona y donde es que se encuentra. Así como poder identificar los puntos particulares del rostro como los son, ojos, boca, nariz, hasta determinar el nivel de atención que las personas emplean en determinada actividad.

Lo anterior, se puede realizar en función de varias señales, como lo son: color de piel (para rostros en imágenes y videos), movimiento (para rostros en videos), forma de la cabeza, apariencia facial o una combinación de estos parámetros [1].

Para la detección de rostros existen diversos algoritmos que permiten eficientemente la identificación de rostros en una determinada escena o una secuencia de las mismas, para lo cual se requiere considerar varios aspectos relacionados con la pose, resolución, ruido, sombra, iluminación, etc.

Mismos de los que se hablarán más adelante para poder llevar a cabo los objetivos de esta tesis, ya que es de vital importancia identificar la teoría asociada a la detección del rostro, la orientación o posicionamiento del mismo con respecto a un punto de observación (cámara web) para poder determinar si está fijando su atención en la escena presentada en pantalla.

1.1. Motivación

Debido a la pandemia mundial por la que estamos atravesando, las tecnologías de la información y comunicación se han convertido en un elemento clave para que las organizaciones, sistemas educativos, médicos, entre otros, funcionen de manera remota, siendo una solución efectiva para limitar la proximidad y el contacto físico, con el objetivo de evitar contagios. Por lo anterior, una alternativa que ha sido ampliamente utilizada por las personas es el trabajo remoto, a través de enlaces digitales (sesiones en línea con video, sin video, plataformas de colaboración, etc.) en los cuales se intercambian datos de interés para los usuarios, en diferentes ámbitos tanto académicos, laborales, etc.

En este sentido, las sesiones basadas en video han permitido la colaboración de las personas en diferentes escenarios. Sin embargo, existen limitantes que se tienen que resolver: por ejemplo, cuando se transmite una información a través del video-enlace, a diferencia de una interacción física, es difícil determinar que las personas están poniendo el respectivo interés en la información presentada en

pantalla. Ya que no se puede deducir de manera efectiva si mínimamente se está mirando a la persona que este presentando.

Por lo anterior, surge la necesidad de medir de forma remota el nivel de atención que un usuario está prestando en una sesión de trabajo. Una alternativa para resolver este problema es el uso de la visión computacional, en la cual, a través del video que puede ser capturado a través de una webcam (las cámaras que tienen los sistemas de cómputo), se detecte a la persona que está trabajando con el sistema de cómputo y midiendo aspectos clave como la orientación que tiene el rostro con respecto a la pantalla del sistema, movimientos que presenta el usuario, movimiento de los ojos, etc., se pueda determinar cuantitativamente el nivel de atención que tiene el usuario.

La motivación principal de este trabajo, tomando en cuenta lo anterior, es la implementación de algoritmos computacionales que permitan identificar si una persona está prestando atención a una tarea de comunicación visual empleando video en tiempo real, podrá ser de apoyo para la toma de decisiones. Lo que permitirá saber que tan bien se está llevando un trabajo en línea, mejorar la calidad de la enseñanza, mantener la coordinación en el trabajo, diversificar las estrategias de aprendizaje y/o transmisión de información según sea el caso.

1.2 Objetivos

1.2.1 Objetivo general

Desarrollar un algoritmo basado en estrategias de visión computacional para determinar si una persona está centrando su atención a lo presentado en pantalla al momento de trabajar con sistemas de cómputo, desarrollando técnicas que permitan detectar indicadores visuales y de posición del rostro para determinar distracción o atención a partir del procesamiento del video en tiempo real.

1.2.2 Objetivos específicos

1. Crear una tabla comparativa analizando las técnicas más relevantes basadas en visión computacional que permitan la detección del rostro y parámetros relevantes para determinar el nivel de atención de una persona, considerando métricas como los tiempos de respuesta, nivel de confiabilidad a partir de la precisión, entre otros.
2. Identificar las variables que permitan conocer el comportamiento “natural” del rostro de una persona al momento de trabajar con sistemas de cómputo por periodos de tiempo prolongados, con respecto a la orientación del rostro, identificación ocular, parpadeo, entre otros.
3. Diseñar un algoritmo que permita identificar mediante la localización de puntos de referencia faciales los movimientos del rostro y parámetros relevantes, que permita analizar si está prestando atención o no.
4. Medir la eficiencia del algoritmo a partir de métricas de precisión acordes al problema planteado.

1.3 Alcances y limitaciones

1.3.1 Alcances

El presente trabajo establece criterios para inferir si un usuario está prestando atención o no a lo que se está presentando en pantalla, lo anterior a partir de la identificación de expresiones faciales como lo son la inclinación, rotación y traslación del rostro, así como el parpadeo y la apertura ocular determinar si está mirando a la pantalla. Por lo que se establecen los criterios que permiten el buen funcionamiento del sistema:

- **Distancia y posición de la cámara.**

La cámara debe de estar a la altura de los ojos, de manera que el rostro del usuario sea visible en su totalidad. El usuario debe estar a una distancia de entre 40 y 60 cm de distancia para que pueda ser procesado de manera

correcta, de lo contrario podría haber pérdida de resolución, ocasionando falsos tanto positivos como negativos.

- **Iluminación**

El usuario debe de estar en condiciones de iluminación aceptables, ni a contra luz, ni sobre expuesto, esto para evitar detecciones incorrectas.

- **Resolución**

La resolución mínima de la cámara web debe de ser de 640 x 480, si la resolución de la cámara es menor pudieran generarse falsos tanto positivos como negativos a la hora del procesamiento de video.

1.3.2 Limitaciones

El trabajo a presentar solo puede deducir a partir de parámetros identificados si el usuario está manteniendo su atención a lo presentado en pantalla, es decir; si está mirando la pantalla, más no determina si el usuario aprende, escucha o entiende la información proporcionada.

Del mismo modo se establece que si los criterios para la correcta identificación no se cumplen pudiera generar falsos a la hora de inferir el nivel de atención ya que pudiera no identificarse el rostro.

- **Distancia**

Si el usuario está demasiado cerca de la pantalla, no será posible detectar claramente el rostro del mismo, por lo cual pudiera generar detecciones incorrectas en la mayoría de los frames.

Lo mismo aplica para cuando el rostro se encuentra alejado del punto de detección.

- **Resolución**

De igual modo si la resolución de la cámara integrada es inferior a la establecida, no se podrá detectar de manera eficiente los parámetros a evaluar por lo que los niveles de atención no serán precisos.

- **Iluminación**

Si el usuario se encuentra al frente de una fuente de luz, esta no permitirá la detección de manera eficiente, por lo que no será posible la detección de manera correcta. Mismo caso para cuando se encuentra en un ambiente poco iluminado.

- **Posición de la cámara**

Si la cámara se encuentra muy arriba de la altura de los ojos o muy por debajo, no podrá detectar los ojos, por tanto no se registraran valores que determinen el nivel de atención.

1.4 Estructura del documento

En el siguiente capítulo, se describen de manera breve los artículos relacionados con el proyecto, dividido en 4 secciones, las cuales incluyen detección de rostro, detección de ojos y parpados, detección de pose, así como el estado del arte sobre la atención.

En el tercer capítulo se detalla el marco teórico empleado para la realización de este proyecto, se describen algunos algoritmos empleados para determinar la atención, la detección y orientación del rostro dentro de una imagen estática o video, técnicas de detección de ojos y parpados, así como la detección de pose de la cabeza.

En el capítulo cuatro, se desarrolla la propuesta de solución, describiendo los algoritmos utilizados en la realización de este proyecto de tesis, así como la manera en que fueron empleados para medir los niveles de atención.

Para el capítulo cinco, se muestran los resultados obtenidos al mediar la efectividad del sistema en cuanto a mediar la atención que el usuario presta a la pantalla se refiere.

En el último capítulo, se presentan las conclusiones generadas a partir del desarrollo del proyecto, así como el trabajo futuro a realizar.

Capítulo II. Estado del arte

El estado del arte en detección de rostros es bastante amplio, se han propuesto diversos enfoques en búsqueda de mejorar los resultados obtenidos por cada autor.

Existe un gran número de áreas en las que se desarrollan aplicaciones para la detección automática de rostros, ya sea como detección de rostros o como paso previo al reconocimiento de ellos, así como de zonas de interés que se estudian dependiendo el problema a resolver.

2.1 Detección de rostros

Con respecto a visión computacional, existen diversos artículos que emplean algoritmos computacionales para diversas tareas tanto de reconocimiento como de detección de rostros. Por lo cual varios investigadores han tomado como base algoritmos existentes modificándolos con el fin de eliminar las desventajas de los mismos, principalmente de rostros en condiciones específicas tales como iluminación, distancia máxima de búsqueda, tamaño de la imagen, variaciones de pose, etc.

En el trabajo realizado por É. K. Shimomoto, A. Kimura y R. Belém [2], presentan un método para la detección de rostros en imagen, encontrando la posición de la cara, los ojos, nariz y boca. Esto lo hacen mediante la segmentación de la piel con un clasificador bayesiano con técnicas de histograma, con el fin de realizar un análisis estadístico del porcentaje de color que existen en los píxeles de la piel, para poder hacer la clasificación de píxeles en una imagen para posteriormente emplear Haar Cascade Classifier para detectar los ojos, nariz y boca, así como de algunas restricciones geométricas asumiendo que el rostro nunca tiene una gran inclinación.

En 2015 los autores Siddiqi, M.H., Ali, R., Khan, A.M., [3] presentan un sistema de Reconocimiento Facial (FER) capaz de proporcionar una alta precisión de reconocimiento, el cual es menos susceptible a la presencia de ruido, como diferentes condiciones de iluminación y variaciones en las características faciales de los sujetos. El sistema utiliza una técnica no supervisada basada en un modelo de contorno activo para la detección y extracción automática de rostros.

En el trabajo "*Two-stage cascade model for unconstrained face detection*" realizado por D. Marčetić, T. Hrkać and S. Ribarić, [4], proponen un modelo en cascada de dos etapas (2SCM) para la detección de rostros sin restricciones, esto a partir de modelos propuestos por otros autores como lo es el Modelo de Piezas Deformable (DMP), el cual se emplea para la detección de rostros, estimación de pose y métodos de localización de puntos de referencia, así como un modelo de Diferencia de Pixeles Normalizados (NDP) para reducir el espacio características pues tiene propiedades anti simétricas, invariantes de escala, entre otras, ambas funcionan con árboles empleando como primer cascada el método NDP y para la segunda etapa el método DMP.

En el artículo presentado por Mutneja, V., Singh, S. [5], realizan una modificación del algoritmo Viola-Jones en combinación con el filtrado de color de piel para reducir el espacio de búsqueda, así como un análisis de funciones a través de AdaBoost para la selección eficiente de clasificadores, además de una aceleración considerable para paralelizar el entrenamiento y la detección de características Haar por medio de GPU.

Logrando resultados que van desde el 87.5% al 99.69% en cuestión de precisión variando las databases empleadas, así como el tamaño de las imágenes de entrenamiento.

2.2 Detección de ojos

Existen diversos artículos en los que se emplean algoritmos computacionales para tareas que tiene que ver con la detección de ojos y parpados. Que van desde detectar el ojo en una imagen o secuencia de ellas, hasta determinar a qué punto está mirando, así como la apertura del ojo, que es frecuentemente utilizada para determinar somnolencia en conductores.

En 2015 [6], Lu, Y.; Zheng, WL; Li, B.; Lu, BL presentan una combinación de movimientos oculares y EEG para el reconocimiento de emociones. En su artículo, adaptan un marco de reconocimiento de emociones multimodal mediante la combinación de movimientos oculares y electroencefalografía (EEG) para mejorar el reconocimiento de emociones. Los resultados de su experimento demuestran que la fusión de modalidades podría mejorar significativamente la precisión del reconocimiento de emociones por medio de los movimientos oculares.

En el artículo "*Face Detection Using Viola Jones Algorithm and Neural Networks*" realizado por Chaudhari et al. [7], Presentan una variación del algoritmo de Viola Jones, en el cual mediante una red neuronal detectan los ojos en el rostro de una persona ya sea usando lentes o sin ellos lo cual permite reducir los falsos negativos, ya que si no detecta ojos o anteojos en una cara, la desecha, en casi el mismo tiempo que el algoritmo original. Logrando una precisión en la detección de rostros del 90%.

Los autores Narayanan Veliyath, Pradipta De, Andrew A. Allen, Charles B. Hodges, and Aniruddha Mitra [8], presentan una investigación de los datos obtenidos a través de un eyetracker, así como se desarrolló un software para recopilar cierta información de las computadoras de los usuarios, se creó una GUI para obtener puntajes auto informados de los usuarios. En la cual detalla cómo los datos obtenidos de un rastreador ocular se pueden usar para predecir la atención de un estudiante como una medida de afecto en el transcurso de una clase. En la cual utilizaron cuatro modelos de aprendizaje automático:

Bosque aleatorio (RF), Máquinas de vectores de soporte (SVM), Aumento adaptativo (AdaBoost) y Aumento de gradiente extremo (XGB), logrando una mayor precisión utilizando la técnica Extreme Gradient Boosting de aprendizaje automático.

2.3 Detección de pose de la cabeza

Para conocer la pose de la cabeza, diversos artículos han empleado combinaciones de algoritmos y técnicas para determinar la posición de la cabeza ya sea en video o en imágenes estáticas.

En 2011, Menchon M [9], trabajo en la “Implementación de Seguimiento de Rostros en Rehabilitación Cognitiva”. En el cual empleo Haarcascade para el reconocimiento de rostro, así como de la bibliotecas facial landmarks para proyectar los punto 2D a 3D y de este modo poder determinar hacia que cuadrante este mirando, partiendo de la distancia calcula entre el usuario y su punto de observación (cámara web), mismo que implementaron a un software para ayudar a pacientes con la rehabilitación y estimulación cognitiva. En el cual muestran datos aceptables que van desde el 90 al 96% de las veces en condiciones de luz aceptables.

En el artículo Visual Analysis of Eye State and Head Pose for Driver Alertness Monitoring [10], presentado en 2013, emplean un Clasificador SVM y un modelo 3D creado en Blender a partir de rasgos faciales 2D, para el monitoreo del estado de alerta del conductor, en una transmisión de video en vivo capturada por una cámara situada al frente del conductor, la cual calcula características visuales como Índice Ocular (EI), Actividad de la pupila (PA) y Postura de la cabeza (HP) para determinar el estado de alerta del conductor.

En el artículo Fine-Grained Head Pose Estimation Without Keypoints presentado en 2018 [11], los autores presentan un manera robusta de determinar la pose

entrenando una red neuronal convolucional para predecir ángulos de Euler a partir de intensidades de imagen a través de clasificación y regresión de poses en bins conjuntos.

2.4 Detección del nivel de atención

Para determinar el nivel de atención que las personas emplean a determinada actividad computacional, se exploraron varios artículos relacionados a inferir información de los usuarios sobre el estado afectivo para lograr determinar la atención.

En el trabajo “Real Time Face Detection And Facial Expression Recognition: Development And Applications To Human Computer Interaction” presentado por M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan [12], desarrollaron un sistema que detecta automáticamente los rostros frontales en transmisiones en vivo, el cual detecta 7 expresiones faciales como lo son neutral, enfado, disgusto, miedo, alegría, tristeza y sorpresa, ya que según sus estudios dichos estados brinda información sobre el estado cognitivo, como el interés, el aburrimiento, la confusión y el estrés, y las señales conversacionales con información sobre el énfasis del habla y la sintaxis. Para lograrlo emplearon Maquinas de Vectores (SVM) para lograr la clasificación de expresiones así como filtros y funciones Haar y Adaboost para la localización de las caras.

En el trabajo presentado por S. Duffner and C. Garcia [13], proponen un método para estimar el foco de atención visual (VFOA) en una secuencia de video capturada por cámara que muestra a varias personas sentadas alrededor de una mesa, la estimación se basa en una Cadena Oculta de Markov (HMM) cuyos parámetros se aprenden de forma incremental, que está integrado en un marco de filtro de partículas globales que se utiliza para el seguimiento facial.

Para 2017 los autores X. Zhang, C. Wu, P. Fournier-Viger, L. Van and Y. Tseng [14], proponen un sistema que integra un módulo de movimiento de la cabeza, un módulo de movimiento del lápiz y un módulo de enfoque visual para analizar con precisión los niveles de atención de los estudiantes en clase. Estos módulos reconocen y rastrean las diferentes actividades de los usuarios, para proporcionar datos de comportamiento al motor de inferencia de atención diseñado, que calcula los niveles de atención de los usuarios. Este motor está equipado con dos nuevos algoritmos de inferencia atención, nombrado el enfoque basado en reglas y el enfoque basado en datos.

Para 2020, los autores N. Krishnnan, S. Ahmed, T. Ganta and G. Jeyakumar, en su artículo “A Video Analytics Based Solution For Detecting The Attention Level Of The Students In Class Rooms” [15] detectan el nivel de atención de los estudiantes por medio del análisis de video, para lo cual emplean una red neuronal de convolución (CNN) para procesar los fotogramas de vídeo clave capturados para determinar somnolencia para deducir si el alumno tiene sueño o no, se ayudan de las expresiones faciales, para percibir el estado psicológico del alumno en el aula, así como de la detección de la mirada se lleva a cabo para examinar si el alumno presta atención a la pizarra.

Capítulo III. Marco teórico

3.1 Reconocimiento facial

El reconocimiento facial es una manera eficiente de identificar o confirmar la identidad de una persona mediante sus rasgos faciales. Los sistemas de reconocimiento facial se pueden usar para identificar a personas desde en fotografías hasta videos o en tiempo real.

El reconocimiento facial es una de las diversas categorías de seguridad biométrica. Otras más avanzadas incluyen el reconocimiento de voz, de huellas digitales y el reconocimiento de retina o iris [16]. La tecnología se utiliza principalmente para la protección y seguridad, aunque ha surgido un interés en otras áreas de uso. Para llevar a cabo este proceso se utiliza la inteligencia artificial, y, en función de las técnicas empleadas, de los recursos disponibles y de los potenciales objetivos, el reconocimiento facial se puede implementar a través de la IA.

El reconocimiento facial tiene múltiples aplicaciones, como lo pueden ser:

- **Seguridad**

Inicio de sesión, desbloqueo de dispositivos móviles y ordenadores, seguridad en aplicaciones, cifrado de información.

- **Vigilancia**

Video vigilancia, control de circuitos cerrados, control de acceso a instalaciones o vehículos, hurtos, seguimiento de sospechosos.

- **Entretenimiento**

Videojuegos, realidad virtual, interacción humano-robot, interacción humano-computadora.

- **Identidad**

Carnet de identificación, carnet de conducir, pasaportes, inmigración, acceso a aeropuertos.

3.2 Algoritmos de detección facial

La detección facial es el primer paso en el reconocimiento facial automático. Este apartado presenta métodos y algoritmos para construir detectores faciales.

3.2.1 Algoritmo de Viola-Jones

El algoritmo de Viola-Jones se presentó en 2001, como un enfoque para la detección de objetos, el cual destaca por su bajo coste computacional, lo que facilita su uso en sistemas que requieren un procesamiento en tiempo real. Viola y Jones presentaron una serie de experimentos detallados en un conjunto de datos de detección de caras en condiciones difíciles de reconocer como: iluminación, escala, variación de cámara y pose. Su proyecto fue movido por el problema de la detección de rostros, donde sigue siendo utilizado muy ampliamente, pero puede aplicarse al reconocimiento de diferentes objetos que estén caracterizados por patrones típicos de iluminación [17].

El algoritmo se basa en una serie de clasificadores débiles denominados Haar-like features que se pueden calcular a partir de la llamada imagen integral [18]. Que son características o rasgos muy simples que se buscan en las imágenes, consisten en la diferenciación de intensidades luminosas entre regiones rectangulares adyacentes, las cuales, quedan definidas por rectángulos y su posición relativa a la ventana de búsqueda que adquieren un valor numérico resultado de la comparación que evalúan.

Estos clasificadores, que por sí mismos tienden a acertar en una probabilidad similar al azar, se agrupan en una cascada, en la cual emplean un algoritmo de aprendizaje basado en AdaBoost para conseguir un rendimiento elevado en la detección de objetos así como una alta capacidad discriminativa en sus primeras fases.

Las principales ventajas que presenta el algoritmo son:

- Su gran velocidad de procesamiento.
- Selección de características eficiente.
- No depende de localización ni escala.
- Trabaja con características, en lugar de pixel a pixel.

El algoritmo tiene cómo desventajas:

- Alta efectividad solo en imágenes de caras frontales.
- No detecta caras que presenten un ángulo de rotación
- Es sensible a condiciones de iluminación.
- Múltiples detecciones de la misma cara.

3.2.2 Detección facial basada en redes neuronales

Una red neuronal se define como un modelo matemático compuesto por elementos de procesado organizados en niveles, estando basada en el funcionamiento de las neuronas biológicas. Esta inspirada en el cerebro humano compuesta por neuronas artificiales, también conocidas como perceptrones, que están conectados entre sí, en múltiples capas. Cada uno consta de una función matemática, ya sea de suma (sumar todas las entradas y pasar) o una función de umbral (limitar la entrada).

En el caso de la detección de rostros, el sistema de red neuronal examina todas las ventanas para determinar si constan de rostros o no. Lo que reduce la tarea computacional ya que no requiere entrenar con imágenes sin rostro.

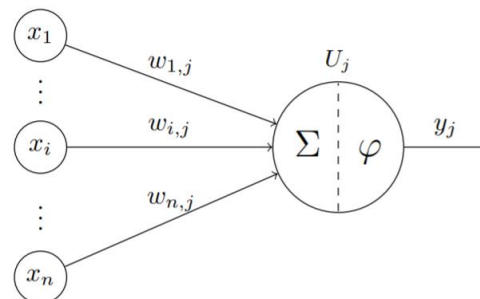


Figura 1. Representación genérica de una neurona artificial

La principal ventaja de las redes neuronales se basa en que su entrenamiento logra como resultado un sistema capaz de capturar e interpretar patrones faciales complejos, su principal desventaja es el tiempo que ha de dedicarse a la creación y ajuste de una arquitectura que permita a la red tener un rendimiento óptimo [19].

3.2.3 Puntos de referencia faciales

Los puntos de referencia del rostro (facial landmarks) se definen como un subconjunto de puntos que enmarcan del rostro, los cuales marcan la punta de la nariz, el mentón, los bordes de los ojos, las comisuras de la boca, las cejas y algunos las orejas. Generalmente en el rostro se reconocen de 60 a 80 puntos, dependiendo la técnica a emplear y los objetivos del estudio a realizar, estos pueden ser tareas de reconocimiento facial, detección de mirada, seguimiento facial, etc.

A continuación se describen de manera breve el funcionamiento de algunas de las bibliotecas de referencia facial.

3.2.3.1 Dlib

Es una biblioteca multipropósito que puede ser usada desde C++ y Python, contiene múltiples algoritmos de aprendizaje máquina, análisis de imágenes, entre otras, esta es una librería de código abierto que debe ser compilada para poder usarla, Dlib cuenta con un kit llamado SVM lineal basado en un histograma de Gradiente Orientado (HOG) para crear detectores de objetos, con herramientas de aprendizaje profundo que permiten la obtención de 68 puntos que enmarcan el rostro así como de las partes relevantes de la cara, ojos, boca, nariz, cejas, cada punto está definido por un conjunto de puntos preestablecidos.

La localización de los 68 puntos para el rostro está codificada dentro de la biblioteca de "imutils" [20].

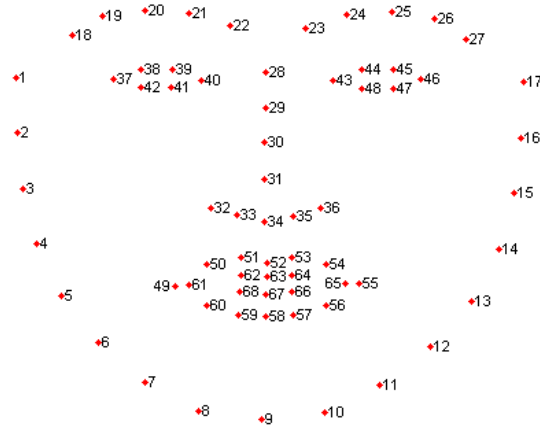


Figura 2. Puntos de referencia faciales presentados por Dlib [20]

3.2.3.2 Flandmark

Flandmark es una biblioteca de código abierto en C (también existe una interfaz para MATLAB y Phyton) que implementa un detector de landmarks en imágenes estáticas. Mediante un clasificador basado en parte en el modelo deformable (DPM), que detecta las coordenadas de puntos clave importantes en rostros frontales. La entrada de esta biblioteca es una imagen con un rostro, el cual devuelve 8 facial landmarks [21].

3.2.4 Extracción de características

Es utilizada para obtener la información más importante del rostro. La cual es posteriormente utilizada para realizar el reconocimiento de la persona. Existen dos tipos de técnicas de este reconocimiento. Las basadas en la apariencia y las basadas en los modelos.

3.2.4.1 Métodos basados en la apariencia (holísticos)

PCA (Principal component Analysis)

Desarrollado por Matthew A. Turk y Alex P. Pentland basado en el principio de análisis de componentes principales de I.T.Jolliffe [24], se basa en un subespacio

en la que las imágenes son representadas basándose en las características más relevantes.

PCA es un método estadístico de análisis de datos y caracterización de los mismos, es una técnica es utilizada para reducir la dimensionalidad de un conjunto de datos, esta puede aplicarse al reconocimiento facial, para ello desarrolla alguna variación llamada Eigenfaces.

SVM (Support Machine Vector)

Se trata de un tipo de clasificadores de patrones basados en técnicas estadísticas propuestas por Vapnik [22] y sus colaboradores en 1992. Es un método genérico para resolver el tema del reconocimiento de patrones. En el cual dado un conjunto de puntos en un determinado espacio que pertenecen a dos clases distintas, SVM encuentra el hiper plano que separa los puntos de la misma clase en el mismo lado. Los puntos más cercanos al hiper plano, en cada conjunto son los llamados vectores de soporte (support vectors).

Es un método que puede ser usado en combinación con otro, ya que este método es discriminatorio, pero incapaz de obtener de las imágenes sus propias características, puede ser empleado con el método Eigenfaces (PCA) para que este pueda extraer las características propias de la imágenes para después emplear SVM para la diferenciación entre las clases, utilizando los valores obtenidos por el método PCA [21].

3.2.4.2 Métodos basados en modelos

HMM (Hidden Markov Models)

Estos modelos propuestos por F. Samaria y S. Young en 1994 [23] utilizan regiones de pixeles que contienen la frente, nariz, boca, ojos, y barbilla, no obtiene la posición exacta de cada rasgo. Este modelo asume que el sistema a modelar es un proceso de Markov de parámetros desconocidos, el objetivo, es determinar los parámetros desconocidos (u ocultos) partir de parámetros observables, es decir, que se conocen.

3.3 Métodos no invasivos para detección de ojos

La detección de ojos, es utilizada en diversas aplicaciones, desde la detección ocular, corneas, etc, con fines médicos, hasta el reconocimiento y autenticación del iris ocular para aplicaciones de seguridad, entre otras. Existen diversos métodos utilizan rastreadores externos para el seguimiento de los ojos.

A continuación se muestran algunas técnicas no invasivas para la detección de ojos, es decir detección desde una cámara web.

3.3.1 Transformada de Hough circular

Es un algoritmo que permite la detección de formas geométricas que pueden describirse como rectas, parábolas, círculos, etc.

Comienza con la detección de bordes para la identificación del contorno, se realiza una búsqueda de círculos que se pueden usar para determinar los parámetros de un círculo cuando se conocen varios puntos que están en el perímetro [25].

La transformada de Hough es una alternativa para la detección de objetos circulares, debido a que el iris es una membrana circular de gran contraste con el ojo, la detección de contornos se puede realizar para aplicarse para detectar la presencia del círculo del iris, incluso si este está cubierto por el parpado.

3.3.2 Haar Cascade

Anteriormente, se presentó cómo detectar rostros utilizando este algoritmo, OpenCV también puede ser utilizado tomando los parámetros necesarios para llevar a cabo la detección efectiva de los ojos. Puesto que existe un archivo XML con los parámetros específicos para detectar ojos en rostros humanos.

3.3.3 Dlib

La región de los ojos también puede ser identificada desde la librería Dlib. Como se mencionó anteriormente Dlib cuenta con un conjunto de 68 puntos de referencia faciales, pudiendo detectar 6 puntos pertenecientes a cada ojo que

devuelve las coordenadas que enmarcan al ojo, lo que hace posible el análisis de la apertura ocular.

3.4 Detección de pose

La estimación de la pose de la cabeza (head pose estimación) es definida como la capacidad para inferir la orientación de la cabeza de una persona con respecto al plano de visión.

Según V. F. Ferrario, C. Sforza, G. Serrao, G. Grassi y E. Mossi en 2002, el rango promedio de movimiento de la cabeza de una persona engloba: una flexión y extensión sagital (movimiento hacia adelante y hacia atrás del cuello) a partir de los 60.4° a 69.6° , una flexión lateral frontal (flexión del cuello de derecha a izquierda) a partir de los 40.9° a 36.3° , y una rotación axial horizontal (rotación de la cabeza de derecha a izquierda) de -79.8° a 75.3° [26].

Generalmente se supone que la cabeza humana puede ser modelada como un objeto rígido. Bajo esto, la cabeza humana se limita a tres grados de libertad, que puede caracterizarse como pitch, roll y yaw, (figura 3).

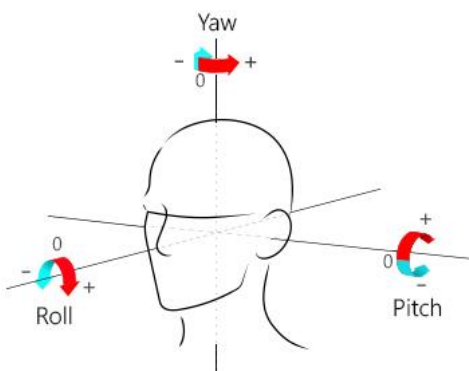


Figura 3. Ángulos de rotación pitch, roll y yaw [26]

La posición de la cabeza ayuda a inferir la mirada, esta puede ser estimada en situaciones en las que los ojos no son visibles (en imágenes de baja resolución, o

en presencia de gafas de sol). Cuando los ojos son visibles, la postura de la cabeza ayuda a predecir con precisión la dirección de la mirada.

3.4.1 Plantillas de apariencia

Estos métodos intentan la localización de la cabeza mediante la comparación de imágenes del rostro previamente entrenadas y etiquetadas con poses. Tiene como ventaja su simplicidad y la capacidad de procesar imágenes de baja resolución, pero solo puede inferir la postura de las poses previamente entrenadas en la plantilla usada.

3.4.2 Conjunto de detectores

Estos métodos se entrenan mediante un clasificador discriminante, como una red neuronal artificial [27], de este modo procesan la imagen de entrada en él. Esta técnica emplea un clasificador binario, previamente entrenado con todas las caras que comparten la misma pose dentro de la base de datos. Esto permite entrenar el detector para distinguir las diferentes posturas de la cabeza, así como la presencia o no de un rostro.

3.4.3 Modelos flexibles

En estas técnicas, un modelo no rígido es superpuesto a la imagen, la cual se ajusta a la estructura facial de cada persona. Para entrenar este método, deben ser localizados y etiquetados manualmente en cada imagen de entrenamiento.

3.4.3.1 Active Appearance Model (AAM)

Aprende de los modos primarios de variación en la forma del rostro y textura desde una perspectiva 2D [28]. Toma un conjunto de puntos faciales específicos M (contorno de los ojos, odios, nariz, la barbilla, boca, etc.), en el que cada punto tiene una coordenada 2D en una imagen y estos son ordenados por características faciales que concatenan en un vector de longitud $2M$. Estos vectores se calculan para muchas caras, que abarcan los diferentes individuos y

las poses de la cabeza, lo que permite que sean utilizados para hallar la variación de la forma de la cara.

3.4.3.2 Active Shape Model (ASM)

Representa los modos primarios de variación de forma [29]. Se basa en los componentes principales, donde se puede encontrar datos de las direcciones que corresponden a la variación en pitch y yaw. La principal limitación es que todas las características faciales tienen que ser localizadas en cada frame de la imagen, esto se limita a las posiciones de la cabeza en la que los bordes de los ojos son detectados. Una ventaja de estos algoritmos es que pueden ser empleados con imágenes faciales de baja resolución.

3.4.4 Proyección 3D

Perspective N-Points también conocida como PNP, consiste en estimar la pose a partir de un conjunto de puntos 3D de coordenadas conocidas, y sus correspondientes proyecciones en la imagen a partir de una cámara.

Tiene como objetivo calcular los diferentes ángulos de rotación de la cabeza con de manera no invasiva, se basa en el seguimiento de las características 2D localizadas en el rostro desde la cámara. Esto se proyecta a un conjunto de puntos de un modelo de cabeza 3D mediante el uso de la perspectiva-n-punto para obtener cabeceo, balanceo y guiñada [30]. Tiene 6 grados de libertad, es decir dos por cada movimiento, donde la posición se expresa como coordenadas cartesianas X, Y y Z.

3.5 Comportamiento frente al ordenador

En general las personas que pasan muchas horas frente a la computadora sufren algunos dolores musculares o de la cabeza, para lo cual la postura es fundamental para evitarlo. Pasar muchas horas trabajando al frente de una pantalla obliga a los ojos a realizar esfuerzos, los cuales pueden producir

molestias oculares, como fatiga visual, ojo seco, picor y visión borrosa que generalmente suelen ser temporales.

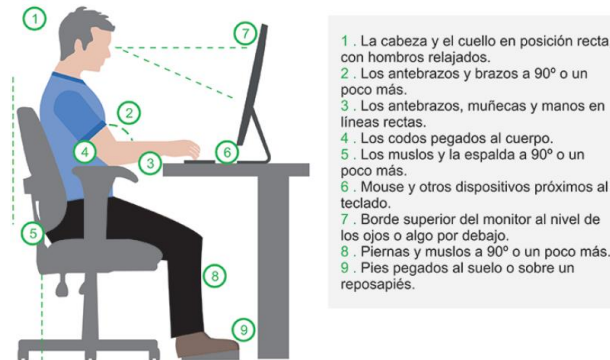


Figura 4. Postura correcta para trabajar frente a la computadora. Tomada de [31]

3.5.1 Distancia entre la pantalla/usuario

La pantalla del equipo de cómputo en el que se trabaja debe de estar por debajo de la línea de los ojos. La distancia recomendada que debe haber entre el monitor y el usuario debe de encontrarse entre 50 y 70 centímetros para que no se forcé tanto la vista [32].

3.5.2 Postura de la cabeza

La postura de la cabeza está determinada por donde este situada la pantalla, ya que este es el punto de observación, la cual debe de colocarse a una distancia, altura e inclinación adecuada.

Además se debería mantener la vista levemente hacia abajo, lo idóneo sería que estuviese a unos 15° por debajo de la línea de mirada. Con el cuello relajado y que el resto del cuerpo tenga una postura no forzada, con la espalda recta.

3.5.3 Iluminación

Normalmente se debería de disponer de una iluminación adecuada, evitando la presencia de deslumbramiento y/o reflejos. La luz que incide en nuestro

ambiente de trabajo debe de estar por arriba, desde el techo si es posible, que sea uniforme, para que no produzca sombras ni deslumbramientos

3.5.4 Parpadeo

El parpadeo de ojos suele ser un acto reflejo mayormente de tipo involuntario. Investigaciones mencionadas en el Libro del ojo seco [33] han determinado que el parpadeo de los ojos desactiva ciertas zonas cerebrales, relacionadas con el sistema visual, permitiendo que, al volver a focalizar lo que se estaba viendo, se le preste mayor atención y se le dé un pequeño descanso al cerebro.

La velocidad con la que parpadean los ojos no siempre la misma y ésta se ajusta a la actividad y en qué momento se esté llevando a cabo. La frecuencia media del parpadeo en canciones de relajación es de 29 por minuto, de 10 cuando se lee un libro y de 7 cuando se trabaja en los ordenadores [34].

3.6 ¿Qué es la atención?

La atención se puede definir como la capacidad de concentrarse y seleccionar estímulos relevantes. En otras palabras, la atención es el proceso cognitivo que permite la orientación hacia los estímulos relevantes y procesarlos para responder en consecuencia [34].

Aleksandr Luria (1962) destacaba como características de la atención la selectividad y la permanencia, consideraba la atención como el factor responsable de extraer los elementos esenciales en la actividad mental, el proceso que mantiene una vigilancia sobre el curso de la actividad mental [35].

Según el psicólogo iraní Albert Mehrabian, profesor en UCLA (Los Ángeles, California), quien es considerado el padre de los estudios sobre el tema, el lenguaje corporal es una comunicación no verbal basada en gestos, posturas, los movimientos del cuerpo y del rostro para transmitir información sobre la persona.

Según el profesor Mehrabian, transmitimos nuestros sentimientos e ideas en una combinación de proporciones que sorprenden: el 7% es verbal, el 38% es vocal (por el tono de voz) y el 55% facial (por nuestras expresiones) [36].

Del mismo modo, la mirada puede transmitir información sobre nuestro estado afectivo, sobre lo que estamos haciendo, la cual puede ser empleada para determinar la atención.

La mirada puede expresar:

- **Interés:** los ojos están despiertos y abiertos, miran tranquilamente y de manera concentrada a la persona que está hablando para absorber toda la información.
- **Desinterés:** la vista se mantiene en constante movimiento.
- **Fatiga:** los ojos parecen más cerrados de lo normal y los parpadeos incrementan considerablemente.

3.6.1 Factores determinantes de la atención

La atención se define como aquella o aquellas variables o situaciones que influyen directamente sobre el funcionamiento de los mecanismos atencionales [37], ya sea para el bien o mal funcionamiento de la misma.

Se presentan los factores más relevantes que determinan la atención:

Motivación e interés

Se fija mayormente la atención en aquello que nos interesa. Por ejemplo, al mirar un artículo, se observan antes los títulos, imágenes o aspectos que sean de nuestro agrado.

Estados emocionales

Cuando las personas son sometidas a estrés o ansiedad, se suele reducir mucho el foco atencional sobre el estímulo que percibimos como amenazante, dejando incluso de atender a otros estímulos.

Estados transitorios

En el caso de los problemas de atención, existen estímulos externos que influyen en la concentración [37]:

- **El ruido.** Se caracteriza por ser estímulo perturbador que suele influir, de forma negativa sobre el nivel de atención de la persona, especialmente cuando la intensidad del ruido es alta o cuando se presenta de forma intermitente. Así mismo, hay ocasiones en las el ruido no muy intenso puede actuar como un activador para que ayude a aumentar los niveles de atención.
- **La falta de sueño.** Uno de los principales efectos de la falta del descanso es el nivel de activación. En consecuencia, disminuye la capacidad para enfocar la atención sobre los estímulos relevantes.
- **Medicamentos.** Afectan a la atención debido a que algunos son tranquilizantes y/o estimulantes. Influye en tareas en las que hay periodos largos que requieren atención, ya que disminuye los niveles de alerta produciendo una baja en la concentración. Los estimulantes ayudan a incrementar el nivel de activación de la persona, por lo que, si no se está muy activo, pueden ayudar a una mayor concentración, teniendo en cuenta que después de esta puede producir lo que es comúnmente llamado como bajón.
- **Sustancias activadoras.** Algunas drogas como lo son la cafeína puede tener efectos positivos en dosis óptimas (400 mgs. al día), ya que mejora los niveles de concentración y aumenta el mantenimiento de la atención. El alcohol, por su parte, tiene efectos negativos sobre la atención pues afecta a la atención focalizada y a la capacidad para oscilar y dividir la atención.

Capítulo IV. Propuesta de solución

En este proyecto se va a emplear un método no invasivo para la detección y análisis del comportamiento del usuario frente a la pantalla de su equipo de cómputo. Para poder determinar el nivel de atención que presta a los eventos de video empleados para diversas actividades.

4.1 Desarrollo de la propuesta

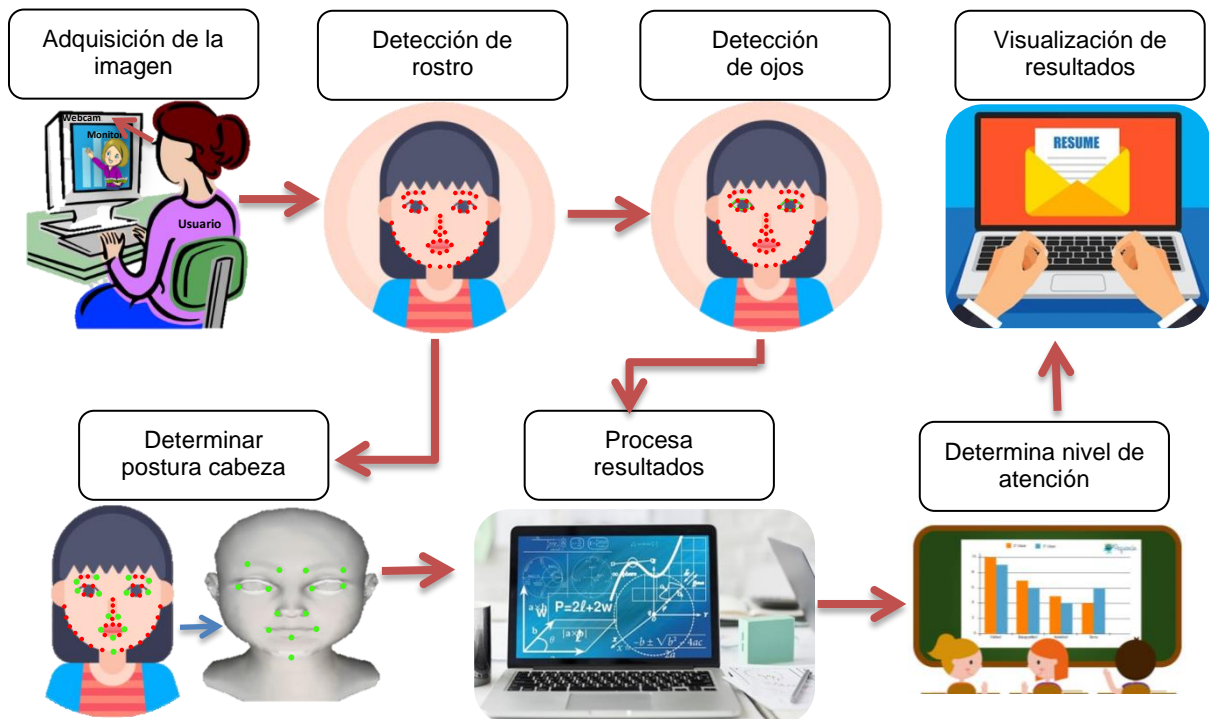


Figura 5. Desarrollo de la propuesta

Las fases a emplear para el desarrollo de la propuesta se describen a continuación:

- **Adquisición de la imagen.**

Obtención de los frames por medio de la cámara web incorporada en el dispositivo de cómputo.

- **Detección de rostro.**

Se detectan los rostros existentes en la imagen por medio de los 68 puntos de referencia facial establecidos por Dlib.

- **Detección de ojos.**

Calculo de EAR, así como conteo de los parpadeos realizados

- **Detección de pose de la cabeza.**

Estimación de la postura de la cabeza por medio del mapeo de los puntos relevantes (explicados más adelante).

- **Procesamiento de los resultados**

Se procesa los frames obtenidos para obtener las regiones de interés y se realizan los cálculos correspondientes en cada frame.

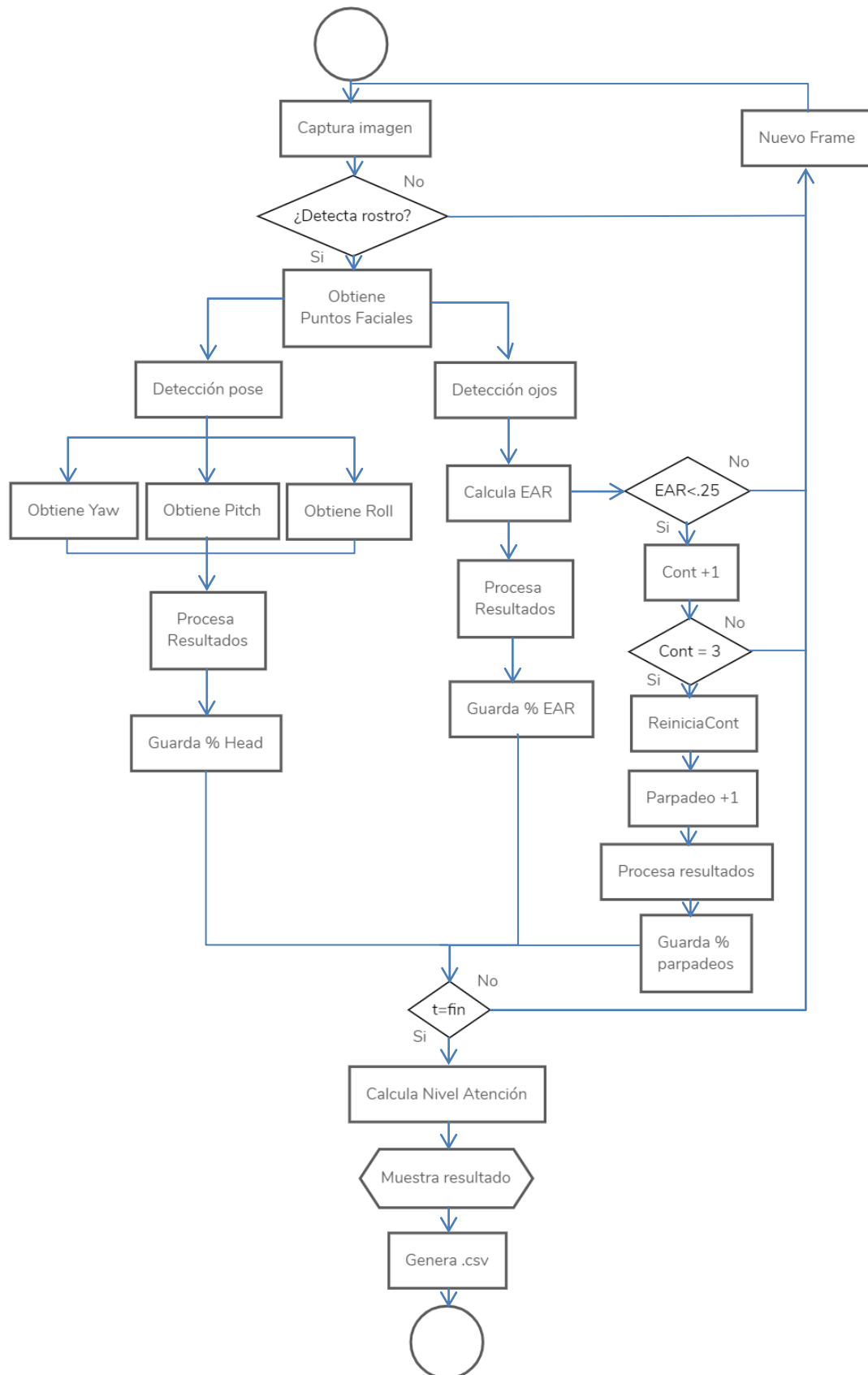
- **Determina el nivel de atención**

Se determina el nivel de atención en base a los resultados obtenidos.

- **Visualización de resultados**

Se muestran el porcentaje de atención presentado durante la sesión de video, así como el tiempo que la persona se mantuvo atenta.

4.2 Diagrama de flujo método propuesto



4.3 Procesamiento frames

Para llevar a cabo el método propuesto, el usuario será grabado a través de su cámara web incorporada en su equipo de trabajo.

En este trabajo, se grabara a 10 fps y se usa una resolución de 640 x 480. Para que las imágenes capturadas por el sistema para la detección de atención sean precisas, se recomienda que el rostro del usuario sea visible en su totalidad, mirando hacia algún punto de la pantalla. Evitando la presencia de deslumbramientos y/o reflejos.

A continuación se describen los métodos empleados para obtener los valores necesarios para determinar la atención, cabe mencionar que estos son obtenidos, procesados y guardados en cada frame.

4.3.1 Detección facial

Para este trabajo se hace uso de la librería Dlib para implementar una función de detección facial. Se localiza el rostro del usuario situado frente a la cámara para posteriormente detectar las distintas marcas faciales dentro de la cara. Dicho método reconoce una serie de puntos faciales que predicen la forma del rostro e intentan localizar 68 puntos de interés que van desde los ojos, cejas, nariz, boca, mandíbula, esto con el fin de evitar errores de detección facial.

Estos puntos son detectados en cada fotograma, una vez detectado el rostro, nos centramos en los ojos, los cuales están representados por 6 puntos cada uno lo cual nos ayudara a hallar la apertura del ojo, así como cuando se produce un parpadeo.

4.3.2 Detección de la relación de aspecto del ojo EAR

Según los autores Soukupová y Cech [38] existe una relación entre el ancho y la altura de las coordenadas oculares, a partir de ello, se deriva una ecuación a la que se le conoce como *Eye Aspect Ratio* (EAR). La relación de aspecto del ojo es una solución eficiente para detectar la amplitud del ojo. Este implica un simple

cálculo basado en la distancia euclidiana. Es decir, en la relación de distancias entre los puntos paralelos de los ojos en vertical y en horizontal.

Mediante un sencillo cálculo se obtiene la apertura del ojo que será empleado para determinar la atención. Ya que cuando miramos hacia la parte superior de la pantalla tenemos los ojos abiertos, lo cual muestra una gran amplitud, ya que al mirar hacia arriba, los ojos se abren considerablemente. Mientras que cuando la mirada va bajando hacia la mitad de la pantalla, la apertura del ojo, por lo general, se encuentra en su estado normal, es decir, no se fuerza el ojo a abrirlo ni a cerrarlo. Por el contrario cuando se mira a algún punto en la parte inferior de la pantalla, la amplitud cae considerablemente.

El cálculo se realiza de la siguiente manera:

$$EAR = \frac{\|p2 - p6\| + \|p3 - p5\|}{2\|p1 - p4\|} \dots\dots\dots [38]$$

Para la determinación de apertura del ojo, se calcula el promedio de la EAR de ambos ojos para reducir errores debidos a guiños o cualquier gesto que implique apertura impar entre los ojos.

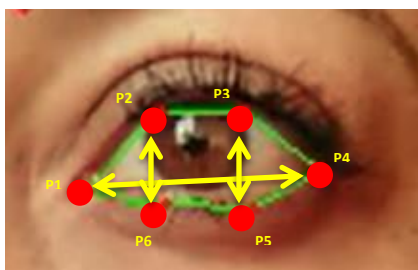


Figura 6. Puntos de referencia ojo izquierdo

4.3.3 Calculo del parpadeo

Por lo general el valor de EAR se mantiene constante cuando se mira al frente y este generalmente tiene un valor alrededor de 25, cuando se produce un parpadeo este valor disminuye llegando a valores por encima de 0.

Debido a que el parpadeo se realiza de manera sincronizada en los ojos, se promedia el valor de EAR de ambos ojos. Trabajos relacionados muestran que el tiempo que tarda una persona en parpadear es, por lo general 300 ms. En este proyecto se obtienen 10 fotogramas por segundo es decir en 1000 ms se capturan 10 fotogramas, siguiendo con lo anterior, si el valor de EAR disminuye a un valor cercano a 0, y se mantiene dicho valor inferior a 25 durante 3 fotogramas seguidos, se considera que el usuario ha producido un parpadeo. De esta manera, se obtiene el número de parpadeos que ha realizado el usuario durante un experimento.

4.3.4 Detección de pose

Para la detección de pose, se toman algunos puntos de referencia facial obtenidos con Dlib y se convierten en puntos a un espacio 3D, para posteriormente pasarlos con *solvePnP* para encontrar vectores de rotación y traslación. Toma 14 puntos del rostro, es decir:

- Ceja izquierda esquina superior izquierda y derecha (17 y 21)
- Ceja derecha esquina superior izquierda y derecha (22 y 26)
- Ojo izquierdo esquina superior izquierda y derecha (36 y 39)
- Ojo derecho esquina superior izquierda y derecha (42 y 45)
- Esquina superior izquierda y derecha de la nariz (31 y 35)
- Esquina superior izquierda y derecha de la boca (48 y 54)
- Esquina inferior del centro de la boca (57)
- Esquina de la barbilla (8)

A partir de las coordenadas 2D, se toman las coordenadas 3D de estos puntos de referencia faciales, así como parámetros internos de la cámara, distorsión de la cámara para estimar los vectores de rotación y de traslación en la punta de la nariz.

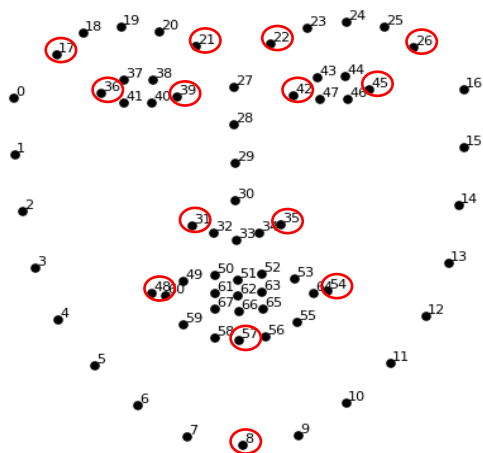


Figura 7. Puntos faciales para detección de pose

Se obtienen para luego calcular la distancia entre los puntos originales y los proyectados con la ayuda de *solvePnP* que resuelve las matrices de rotación y traslación tomando como parámetros *object_pts* que corresponde al sistema de coordenadas mundiales (UVM), *image_pts* que corresponde a los puntos de referencia 2D, *cam_matrix* corresponde a la matriz K y *dist_coeffs* correspondiente a la matriz D.

Posteriormente convertir la matriz de rotación en un vector de rotación para lograr calcular el ángulo de Euler.

4.4 Procesamiento de los resultados

Con los valores obtenidos se hace un análisis para determinar el nivel de atención, el cual consiste en verificar si el usuario estuvo manteniendo los movimientos establecidos como “normales” mientras estuvo presente en un enlace de video. Recordando que en este proyecto trabajamos a 10 fps, los valores obtenidos son procesados en cada frame, por lo cual se obtienen 10 resultados por segundo.

Al finalizar el tiempo establecido para la sesión de video el programa termina mostrando en una ventana el último frame capturado indicando el promedio de

nivel de atención obtenido, así como el tiempo que el usuario mantuvo su atención por arriba del 70%. Siendo este un umbral establecido para esta tarea en específico. Del mismo modo genera un archivo csv con los valores obtenidos en cada fotograma, el cual consta de 7 columnas en las cuales muestra el porcentaje de cada variable, así como el indicativo del fotograma en el cual se obtuvo dicho valor y el nivel de atención en cada uno.

Frame	Pitch	Yaw	Roll	Blink	EAR	LevelAttention
0.10	67.062	68.132	83.974	100	73.751	58.585
0.20	77.723	72.164	82.392	100	87.307	66.203
0.30	74.569	82.639	81.012	100	65.239	57.263
0.40	80.650	79.792	83.256	100	61.463	56.040
0.50	70.657	81.941	80.601	100	97.145	71.069
0.60	69.416	80.661	81.928	100	99.459	72.184
0.70	74.730	91.134	83.274	100	91.243	70.408
0.80	71.786	86.815	82.774	100	95.039	71.150
0.90	73.846	86.999	81.912	100	90.225	69.136

Tabla 1. Ejemplo de los resultados obtenidos a la finalización del sistema.

En la tabla 1, se muestran parte de los resultados que el sistema arrojo, los valores que se muestran, son los generados durante un segundo.

4.4.1 Relación de aspecto ocular (EAR)

La relación de aspecto ocular varia, yendo desde 0 (cuando el ojo se encuentra cerrado) hasta los 0.40 cuando el ojo se encuentra abierto, por lo que se establece un umbral de apertura del 0.35 tomando en cuenta que el tamaño de los ojos varia de persona a persona, por lo cual si EAR se mantiene en dicho umbral o es mayor se considera que se está poniendo el 100% de atención a lo presentado en la computadora. La EAR puede moverse entre los 0.25 a los 0.35 para considerarse una apertura normal, el cual hablando de porcentaje seria de un 71% al 100% de apertura ocular, tomando un promedio del 85%.

Ya que la apertura ocular tiende a ser menor cuando se comienzan a presentar estados de somnolencia. Por lo anterior entre menor sea la relación de aspecto ocular, menor será su indicativo de atención con respecto a la apertura ocular.

4.4.2 Parpadeo

De acuerdo a la bibliografía consultada, los parpadeos por minuto de una persona mientras se encuentra trabajando en un equipo de cómputo se reducen a la mitad, ya que la atención se concentra tanto en la pantalla que olvidamos parpadear de manera frecuente. Por lo que el umbral establecido para los parpadeos va desde los 5 hasta los 10 parpadeos por minuto. Si el usuario permanece en estos valores, se considera que está poniendo el 100% de atención con respecto a los parpadeos realizados, si se rebaza de dicho umbral, el nivel de atención ira disminuyendo conforme más parpadeos se realicen.

De igual manera si el usuario está por debajo de lo establecido se debiera pensar que la persona a detectar esta estática, cosa que es prácticamente imposible al estar en frente al computador por un largo tiempo.

4.4.3 Movimientos de la cabeza

En los trabajos realizados por los autores consultados, establecen umbrales de movimiento en los cuales el rostro de la persona puede ser detectado de manera eficiente va desde los 0 a los 30 grados en sus tres ejes de movilidad. Para los cuales se establece un umbral de 25 grados para yaw, pitch y roll.

Si los valores obtenidos rebasan dichos márgenes es un indicativo de que el usuario dejo de mirar la pantalla por lo cual ha dejado de poner atención, lo cual hace que su nivel de atención marque cero. Por el contrario si la persona mantiene su cabeza derecha, permitiendo movimientos hasta los 5 grados, su nivel de atención estará por el 100%. Los grados de movilidad para inferir que se mira a la pantalla son hasta los 21 grados, lo cual pertenece a un 70% de atención.

4.5 Determinar nivel de atención

Para determinar el nivel de atención final del usuario, se conjuntan los resultados obtenidos en cada apartado (pose cabeza, apertura ocular, número de parpadeos) y se les asigna un peso para poder determinar su atención.

4.5.1 Nivel de atención por pose de la cabeza

- **Yaw.**

A este movimiento se le asigna un peso del 40%, ya que si el usuario sale del umbral establecido, claramente no está mirando la pantalla, por tanto está perdiendo el interés de lo presentado en su computadora.

$$\text{Atten_Yaw} = \text{yaw} * 0.40$$

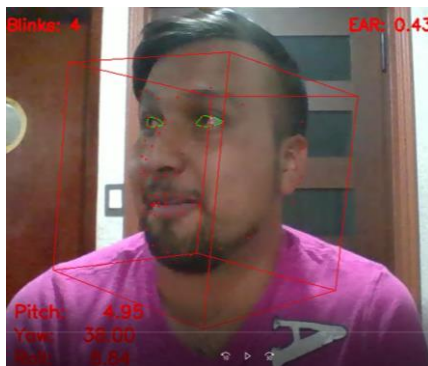


Figura 8. Usuario mirando fuera de la pantalla. Yaw= 38°

- **Pitch.**

Al movimiento denominado como cabeceo, se le asigna un peso del 35%, debido a que los movimientos hacia abajo generalmente son para mirar el teclado, por lo contrario si el movimiento es hacia arriba se puede inferir que mira por arriba de la pantalla, perdiendo la concentración de lo presentado.

$$\text{Atten_Pitch} = \text{Pitch} * 0.35$$

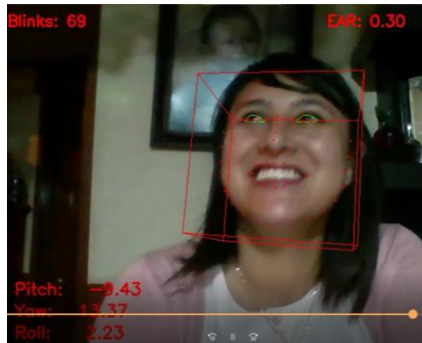


Figura 9. Usuario mirando por encima de la pantalla. Pitch= -9.43°

- **Roll.**

A la flexión lateral que se realiza mientras miramos trabajamos en la computadora, se le asigna un peso de 25%, ya que este movimiento es más flexible respecto a que podemos seguir mirando la pantalla, aunque este un poco inclinada nuestra cabeza.

$$\text{Atten_Pitch} = \text{Roll} * 0.25$$

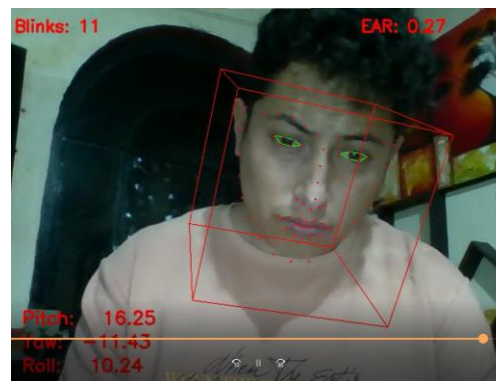


Figura 10. Usuario mirando a la pantalla con flexión lateral. Roll=10.24°

4.5.1.1 Postura final cabeza

Para obtener un solo nivel de atención de la postura de la cabeza. Se suman los valores obtenidos en cada ángulo de movimiento al que le han sido asignados los pesos. Posteriormente se le asigna un peso a dicho movimiento para poder tener una comparación con los demás variables medidas.

$$\text{levelhead} = \text{Atten_Yaw} + \text{Atten_Pitch} + \text{Atten_Roll}$$

$$\text{levelhead} = \text{levelhead} * 0.35$$

4.5.2 Nivel de atención por EAR

La apertura ocular es muy importante para nuestro estudio ya que nos ayuda a inferir si una persona está cansada o soñolienta, puesto que su apertura ocular se mantiene baja en dichas circunstancias. Del mismo modo la relación de aspecto ocular varía de persona a persona dependiendo en su mayoría del tamaño de sus ojos. Tomando en cuenta lo anterior, se le asigna un peso 35%.

$$\text{levelear} = \text{ear} * 0.35$$

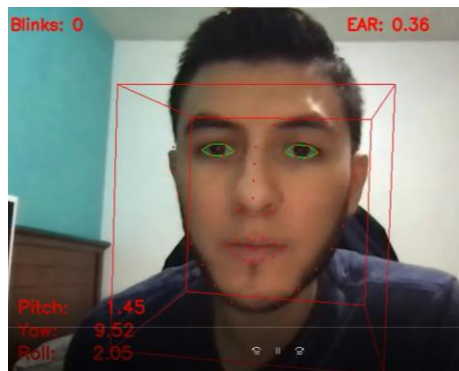


Figura 11. Apertura ocular máxima del usuario. EAR=0.36

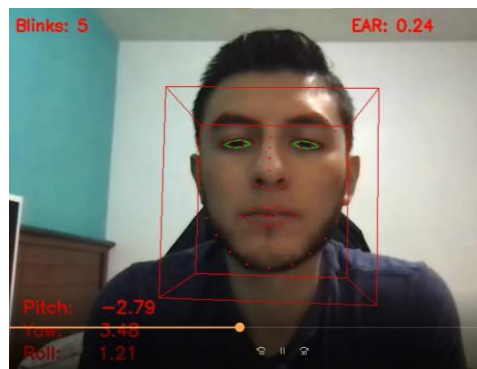


Figura 12. Apertura ocular mínima del usuario, sin producir parpadeo. EAR=0.24

4.5.3 Nivel de atención por parpadeo

Como se ha mencionado el parpadeo es un movimiento involuntario que se realiza para humectar nuestros ojos, dicho reflejo varía de acuerdo a las actividades realizadas, por lo que se le ha asignado un peso del 30% del total de atención.

$$\text{levelBlink} = \text{blink} * 0.30$$

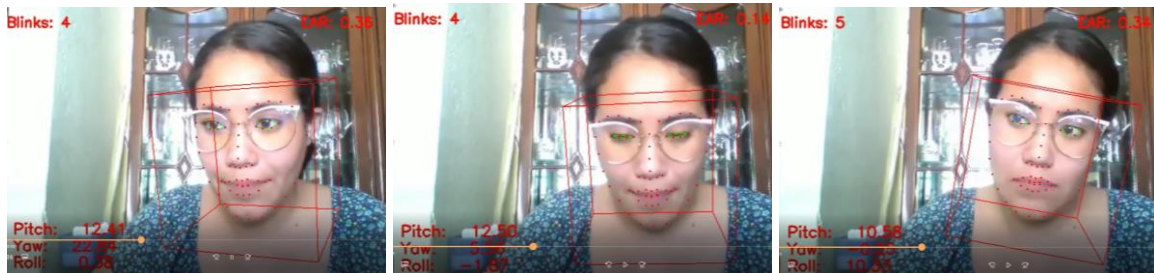


Figura 13. Apertura ocular antes, durante y después de un parpadeo

4.6 Desarrollo de fase experimental

A continuación se describe el proceso seguido para el desarrollo del trabajo presentado. Se describen las especificaciones técnicas y herramientas utilizadas. Se presenta el procedimiento seguido y las fases de desarrollo.

4.6.1 Software

El entorno de desarrollo empleado fue Visual Studio Code, el cual ha resultado muy útil y cómodo para la implementación del código de programación en Python. En el cual se ha creado un entorno virtual de trabajo para todo el proyecto, en el cual se han instalado las librerías que se han utilizado en dicho entorno de trabajo, evitando así interferencias con otras librerías o archivos externos de otros posibles trabajos. En la Tabla 2 se muestran el software empleado para la realización del proyecto.

Software	Descripción
Visual Studio Code 1.65.1	Entorno de desarrollo
Python 3.7.0	Lenguaje de programación
Excel	Hoja de cálculo y análisis de datos
Word	Edición de documento

Tabla 2. Software empleado

4.6.2 Hardware

En lo referente al hardware utilizado, solamente se ha utilizado una computadora portátil (con WebCam) para la realización de todo el proyecto, en la tabla 3 se describirán sus especificaciones técnicas.

Especificación	Propiedades
Modelo	Lenovo 2019 Flex
Procesador	Intel Core i7-8565U CPU @1.80GHz 1.99 GHz.
Tipo de Sistema	64 bits
Sistema operativo	Windows 10 Home
Memoria RAM	8 GB
Pantalla	15.6" HD (1920 x 1080)
WebCam	HD (1280 x 720) en video estándar a una velocidad máxima de 30 fps.

Tabla 3. Hardware utilizado

4.6.3 Librerías de Python

A continuación se enlistan librerías de Python utilizadas. Cabe mencionar que todas las librerías se han instalado mediante el Terminal de VSC.

OpenCV (Open Source Computer Vision Library)

Librería multiplataforma de software de visión por computador y machine learning que incorpora más de 2500 algoritmos optimizados, los cuales se pueden

usar para detección y reconocimiento de caras, rastrear movimientos en video, identificar objetos, clasificar acciones humanas, y un largo etcétera. Se trata de una librería de código abierto desarrollada en 1999 por Intel bajo licencia BSD, por lo que se puede usar y modificar el software libremente, siempre y cuando no se elimine el aviso de copyright.

En la tabla 4, se muestran las librerías que fueron utilizadas para la realización de este proyecto, así como la versión de cada paquete.

Librería	Versión	Descripción
Dlib	19.22.0	Aprendizaje automático
numpy	1.21.5	Procesamiento matemático
SciPy	1.6.0	Procesamiento matemático
Imutils	0.5.3	Procesamiento de imágenes
Pandas	1.3.5	Manejo y análisis de estructuras de datos
tkinter	8.6.10	Interfaces graficas

Tabla 4. Librerías empleadas en el proyecto

Capítulo V. Resultados

En este apartado se presentan los resultados tanto cuantitativos como cualitativos, obtenidos en diversas pruebas realizadas a 20 personas diferentes, entre los cuales se estudió el comportamiento frente a la computadora de 12 hombres, 1 niño y 7 mujeres de entre 12, 20 y 40 años respectivamente. Se analizaron los movimientos de los usuarios en diversas tareas de cómputo para obtener el nivel de atención prestado en cada tarea computacional, así como en diferentes escenarios para verificar la funcionalidad del sistema.

Se les pidió a los usuarios realizar una sencilla instalación del sistema, indicando las librerías necesarias, así como las versiones tanto de python como de la paquetería empleada.

Se les solicitó a los usuarios sentarse frente al ordenador a una distancia de entre 40 y 60 cm de la pantalla. Además, la cámara web de la computadora se situó a la altura de los ojos con el fin de obtener una lectura correcta de la imagen.

Fueron solicitadas 4 tareas que se realizan normalmente en los equipos de cómputo tanto en el área académica como laboral. Con fines de verificar la funcionalidad del sistema, los usuarios fueron grabados durante 4 minutos por cada prueba. Pidiendo las siguientes actividades:

- Tarea 1: Clase o reunión en línea
- Tarea 2: Lectura desde la computadora
- Tarea 3: Visualización de video, redes sociales o video Juegos
- Tarea 4: Escritura de un documento

5.1 Reporte de usuarios/ nivel de atención

A continuación se presentan los comentarios que realizaron los usuarios al momento de presentar su prueba, así como el nivel de atención marcado por el sistema.

5.1.1 Tarea 1. Clase en línea

El usuario 9 indica que al momento de realizar su prueba número uno, se encontraba en una clase que estaba teniendo de su maestría, para la cual se mantuvo atento pero comenta que se sentía un poco cansado ya que se había estado desvelando con anterioridad.

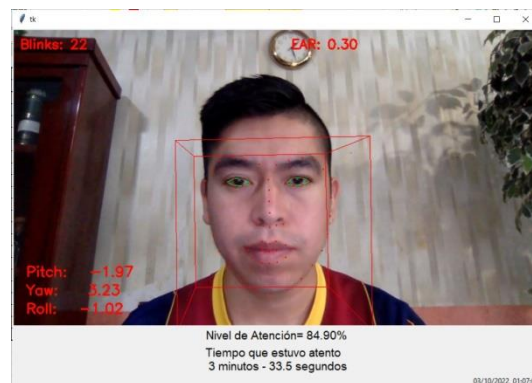


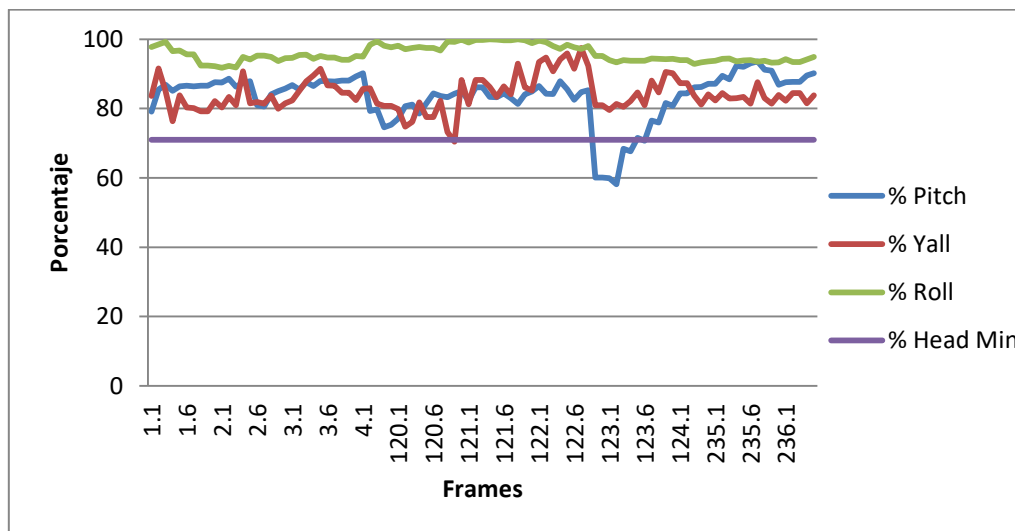
Figura 14. Usuario 9 en tarea 1

Según los resultados obtenidos su nivel de atención se encontró casi por el 85%, su tiempo de atención fue de 3 minutos, con 33 segundos, pudiendo decir que el usuario se mantuvo atento durante su clase.

Con respecto a los 22 parpadeos realizados por el usuario, estos se mantuvieron dentro de los márgenes establecidos, resaltando que tiene el 100% de atención en esta variable.

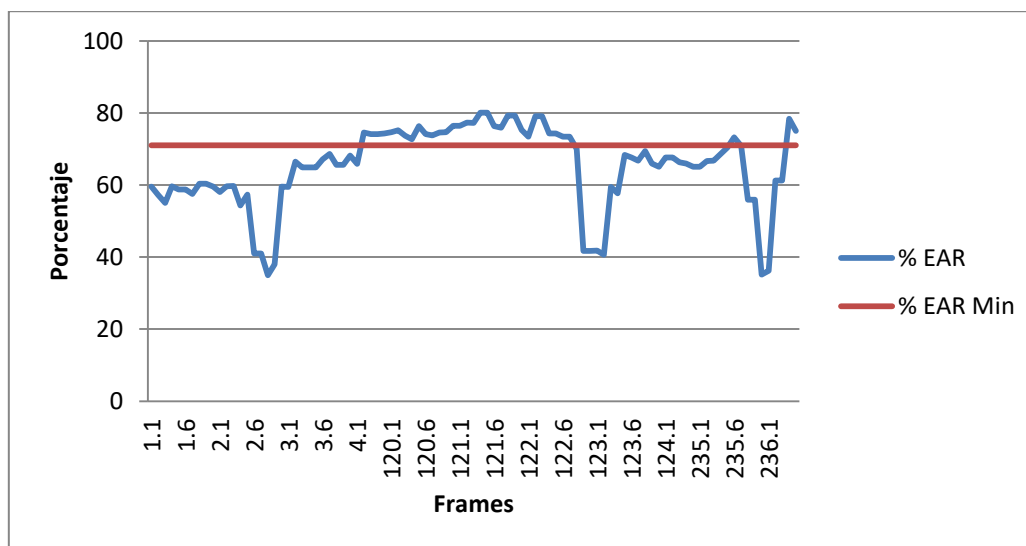
En la gráfica 1, se puede observar los porcentajes obtenidos durante los segundos iniciales, intermedios y finales que fueron capturados por el sistema, en el que se puede apreciar que sus movimientos en la cabeza tanto en Pitch, Yaw y Roll se

encuentran por encima del mínimo, notando la posición de su cabeza se mantuvo atenta durante la mayoría de los frames.



Gráfica 1. Porcentajes de la cabeza obtenidos por el usuario 19

La relación de aspecto ocular del usuario, se muestra en la gráfica 2, como se puede observar su EAR es bajo en la mayoría de los frames, su porcentaje de apertura ocular es del 71.46%, pudiendo inferir que como el usuario lo reporta se encontraba cansado y esto ocasiono que su porcentaje de relación de aspecto ocular fuera menor, bajando su nivel de atención.



Gráfica 2. EAR del usuario 9, durante la tarea 3

5.1.2 Tarea 2. Lectura de documento academico

El usuario 1, comenta que al momento de realizar la prueba se encontraba leyendo un articulo sobre la reforma fiscal, requerido para realizar un trabajo de la escuela, indica que se sentia con la vista cansada debido a que por su trabajo pasa muchas horas frente a la computadora.

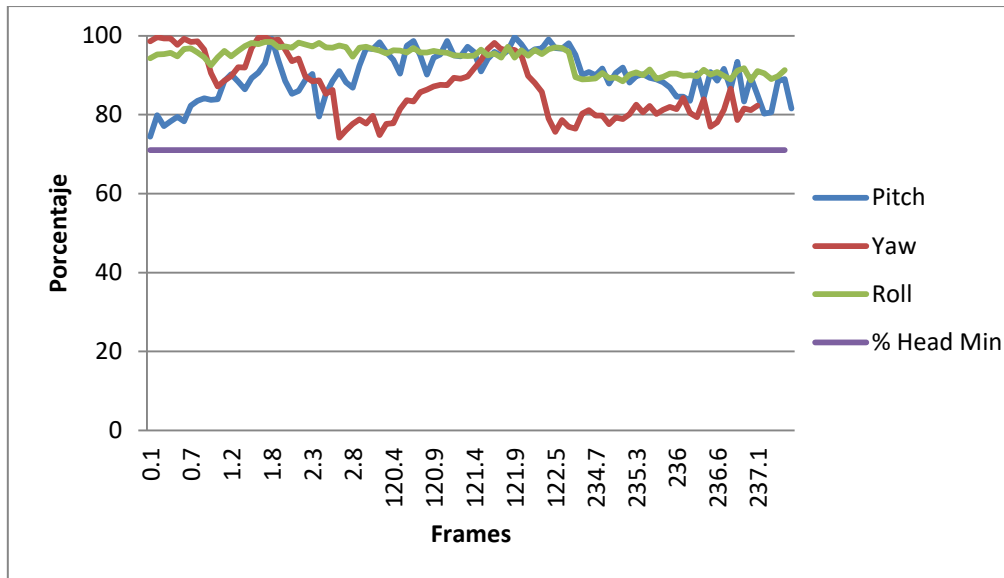


Figura 15. Resultados de usuario 1

El sistema indica que el usuario presenta un 83.50% de atención, del mismo modo muestra que estuvo atento 3 minutos con 9 segundos, pudiendo decir que este se mantuvo atento a su lectura académica.

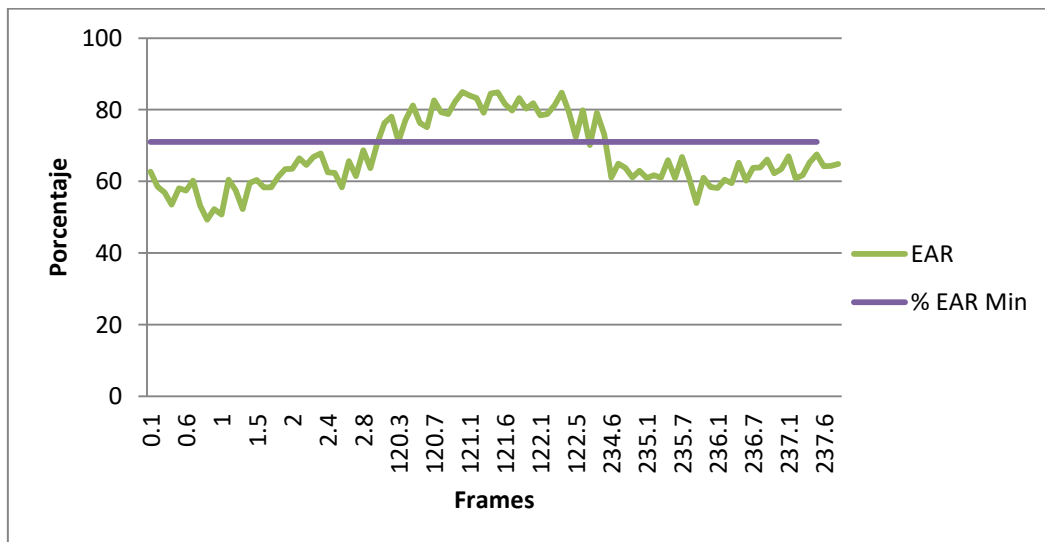
Sus parpadeos se encontraron dentro del rango establecido, indicando un 100% de atención con respecto a los parpadeos realizados.

En la gráfica 3, se muestran los resultados obtenidos de 15 segundos del video, mostrando que los movimientos de su cabeza se encontraban variantes pero por encima del mínimo para inferir que se mira hacia la pantalla. Su movimiento promedio de la cabeza es del 87.27%



Gráfica 3. Movimientos de la cabeza del usuario 1

Los resultados de algunos frames de EAR del usuario, se muestran en la gráfica 4, en la cual la mayoría de sus frames capturados con respecto a su apertura ocular se mantiene por debajo del mínimo establecido, esto podría deberse al cansancio referido por el usuario.



Gráfica 4. Variaciones EAR usuario en tarea 2.

5.1.3 Tarea 3. Actividad de interes

El usuario 19, menciona que durante la prueba número 3, se encontraba viendo videos sobre maquillaje, reporta que estuvo atenta a lo que estaba viendo en la pantalla.

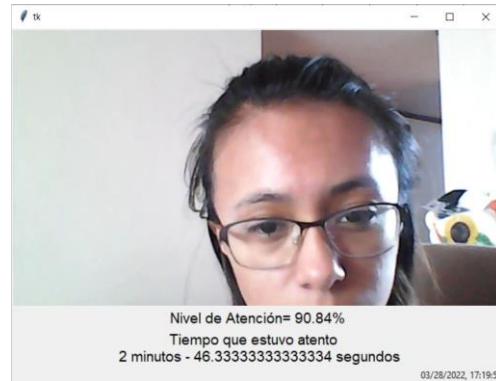


Figura 16. Resultados del usuario 19.

De acuerdo a los resultados obtenidos, en efecto la usuaria estuvo prestando atención a la pantalla, mostrando un nivel de atención por arriba del 90 por ciento, pero debido a que se encontraba demasiado cerca de la pantalla, no fue posible detectar su rostro en algunos fotogramas, ocasionando que su tiempo de atención baje, debido a que este resultado se obtiene de los frames capturados en los que se puede identificar el rostro.

Sus parpadeos se encontraron en 20, los cuales se encuentran dentro del umbral establecido, por tanto su nivel de parpadeo se encuentra en el 100%.

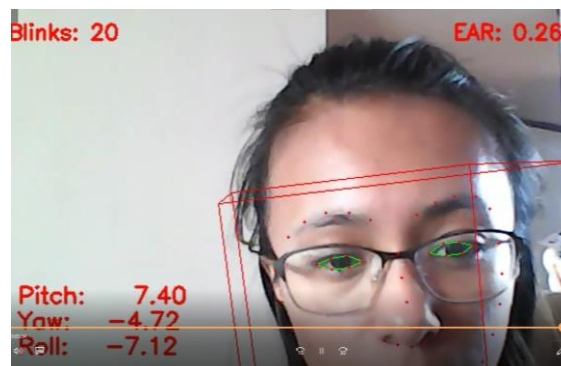
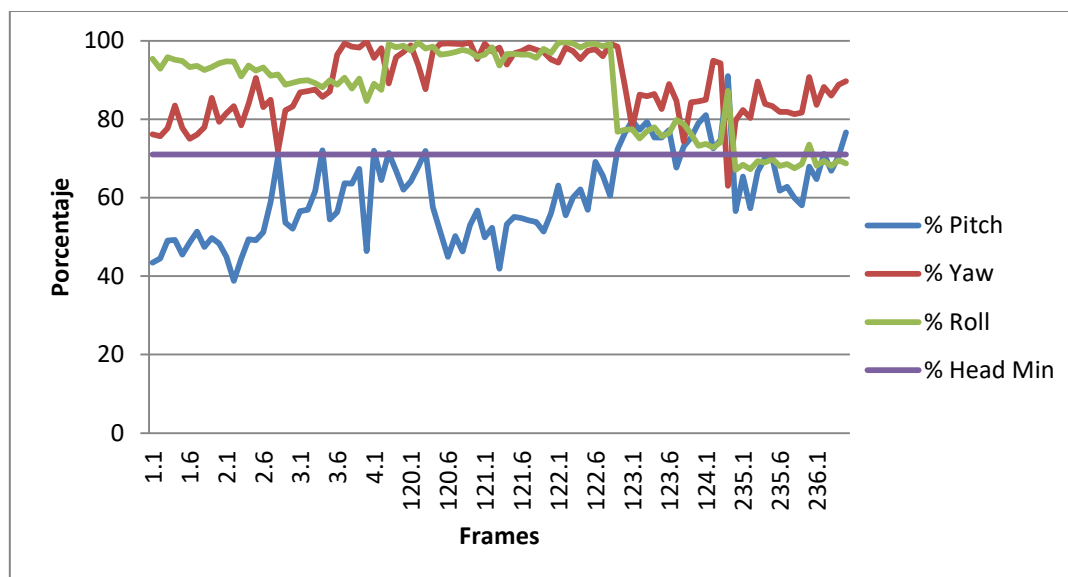


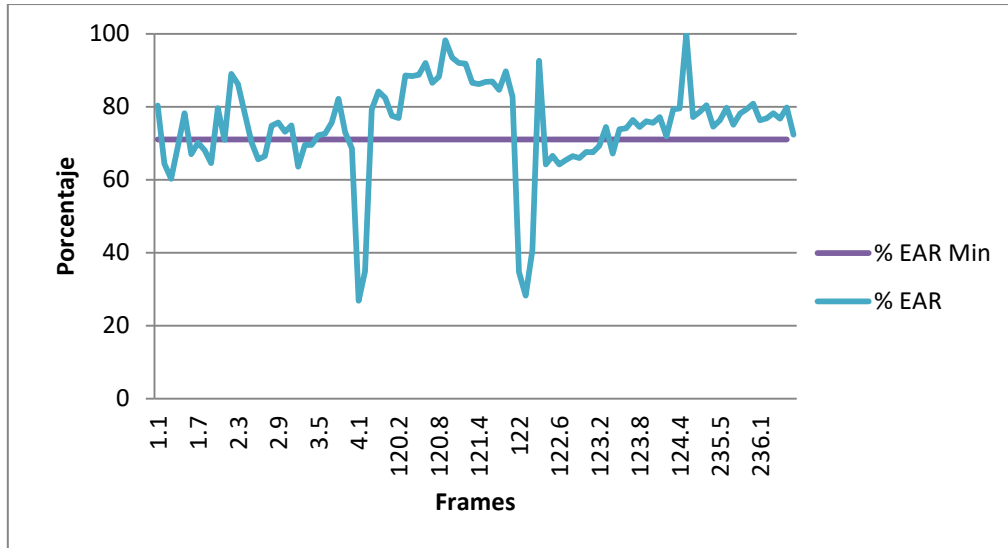
Figura 17. Parpadeos totales realizados por el usuario

En la gráfica 5, con fines de poder apreciar los movimientos de la cabeza, se muestran los primeros 5 segundos del video, 5 segundos intermedios y los últimos 5 segundos correspondientes a los frames capturados. Como se puede observar, el movimiento en Yaw y Roll, se mantiene por encima del mínimo establecido para inferir que mira hacia la pantalla, sus movimientos en Pitch, se mantiene en la mayoría de los frames por debajo del umbral mínimo, esto debido a que su cámara se encontraba mal posicionada, ya que estaba demasiado cerca y por encima, por lo que su rostro no era totalmente detectado, por lo que sus movimientos no eran claramente identificados.



Gráfica 5. Movimientos de la cabeza realizados por el usuario

En la gráfica 6, se muestra la relación de aspecto ocular de 15 segundos, tomando 5 segundos iniciales, intermedios y finales, su EAR promedio es de 78.19%, pudiendo notar que este se mantiene cerca del mínimo, tanto por arriba como por debajo, del mismo modo en la gráfica se observa cuando fue producido un parpadeo.



Gráfica 6. EAR presentado por el usuario

Respecto a su tiempo de atención, se encontró en 2 minutos con 46 segundos, los cuales como ya se ha mencionado variaron debido a que algunos frames no fueron capturados por el posicionamiento y cercanía de su rostro.

5.1.4 Tarea 4. Escritura de documento

Para la cuarta tarea, el usuario 17 mensiona que realizo su prueba cuando se encontraba redactando un artículo académico.

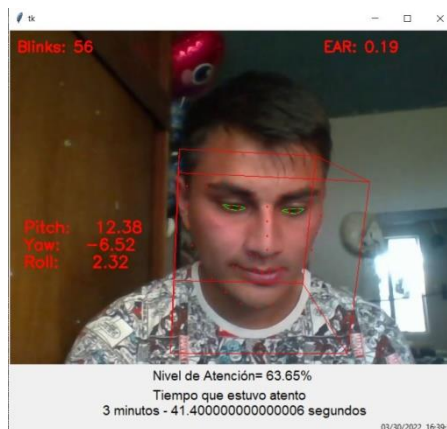
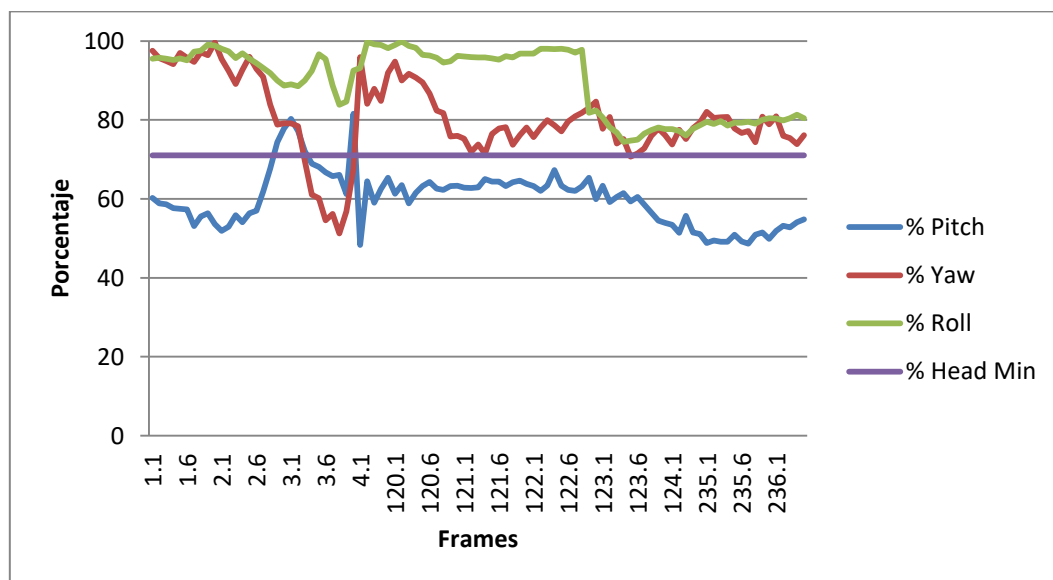


Figura 18. Resultados obtenidos en actividad de escritura

El nivel de atención del usuario definido por el sistema es del 63.65%, así como un tiempo de 3 minutos con 41 segundos. Lo cual nos indica que el usuario estuvo atento la mayor parte del tiempo, pero debido al movimiento de su cabeza, es decir agachar la cabeza para mirar el teclado, como del EAR en el que sus ojos tienen menor apertura al mirar el teclado como de los parpadeos que se producen más debido a que permanece con los ojos por debajo del umbral varios frames consecutivos, su nivel de atención baja considerablemente.

Al final de la prueba sus parpadeos se registran en 56, es decir 16 parpadeos más de los establecidos como normales.

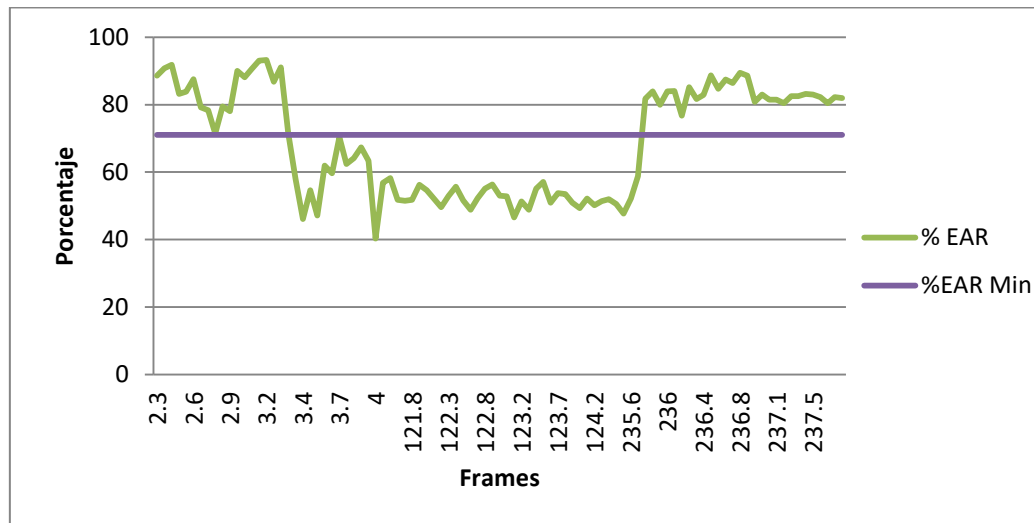
En la gráfica 7, se muestran 15 segundos tomados de los frames capturados al inicio a la mitad y al final de la prueba. Como se puede observar el movimiento de Pitch es el que se encuentra por debajo del establecido para indicar que se mira la pantalla, en el caso de esta tarea esta variación es normal ya que el usuario bajaba su cabeza para mirar el teclado.



Gráfica 7. Movimientos de la cabeza usuario 17

La relación de aspecto ocular del usuario se muestra en la gráfica 8, pudiendo notar que su apertura baja cuando se miraba hacia el teclado.

Se EAR promedio fue del 60.82%



Gráfica 8. Relación de aspecto ocular del usuario 17

5.2 Resultados obtenidos

Se muestran los resultados obtenidos en las 20 pruebas realizadas a los diferentes usuarios a quienes se les pidió usar el sistema. Se muestra los resultados arrojados por el programa en cada tarea.

Se señalan las pruebas en las cuales los usuarios se encontraban demasiado cerca, mal posicionados, con baja resolución o con mala iluminación ya sea con poca o a contraluz.

5.2.1 Tarea 1

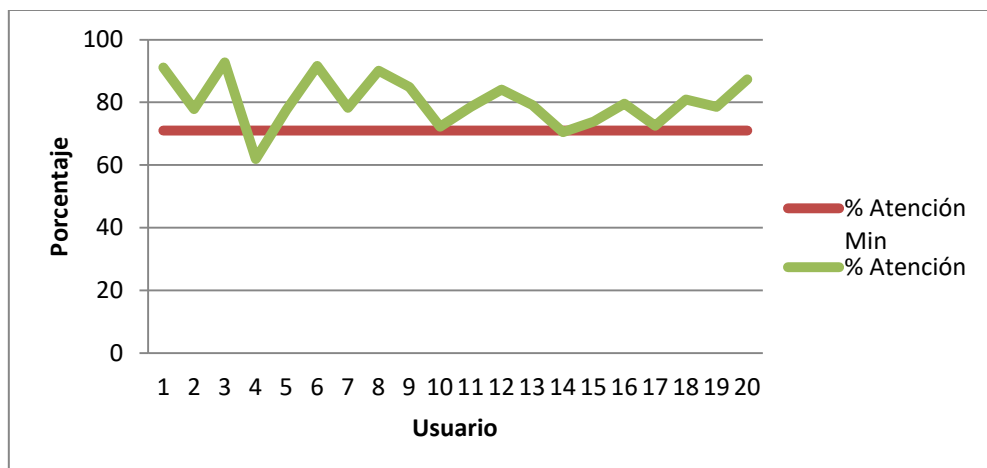
En la siguiente tabla se muestran los resultados del porcentaje de nivel de atención así como del tiempo en el que los usuarios se mantuvieron atentos expresado en minutos y segundos.

Usuario	Tarea 1		Usuario	Tarea 1	
	% Atención	Tiempo		% Atención	Tiempo
1	91.16%	3 min 44s	11	78.51%	2 min 42s
2	77.87%	3 min 09s	12	84.05%	3 min 55s
3	92.71%	3 min 16s	13	79.14%	3 min 50s
4	61.91%	3 min 10s	14	70.50%	3 min 29s
5	77.60%	2 min 43s	15	73.86%	3 min 13s
6	91.58%	3 min 18s	16	79.56%	2 min 57s
7	78.22%	2 min 27s	17	72.56%	3 min 17s
8	90.06%	3 min 42s	18	80.90%	2 min 44s
9	84.90%	3 min 33s	19	78.57%	2 min 03s
10	72.19%	3 min 28s	20	87.33%	3 min 37s

Tabla 5. Resultados de tarea 1

En la tabla 5 se muestran los resultados obtenidos en la prueba. En la cual se les solicito grabarse mientras se encontraban en clase en línea o tenían alguna reunión del trabajo.

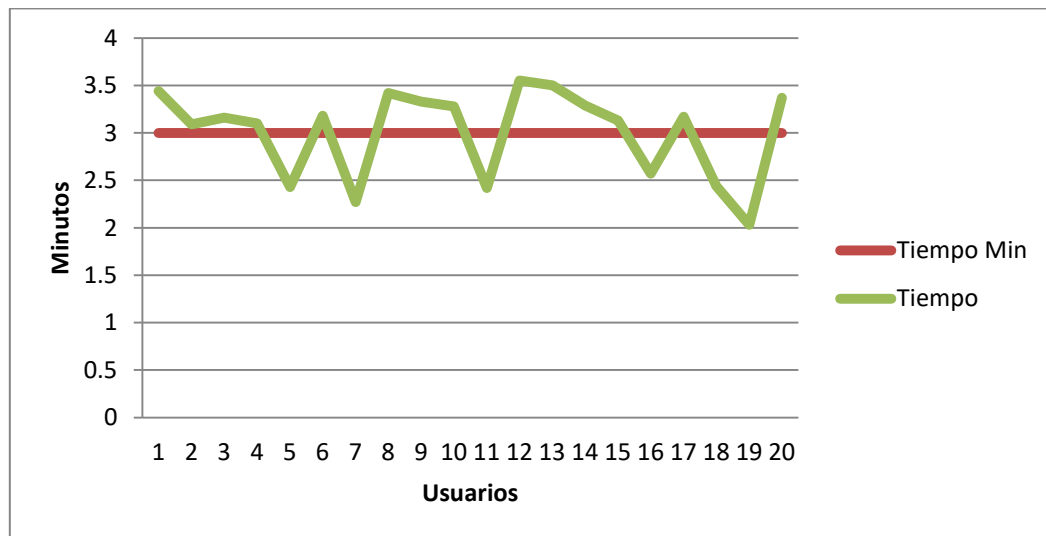
Como se puede observar en la gráfica 9, el 90% de usuarios presta atención a lo que estaban visualizando en pantalla, se refiere que solo 1 usuario se encontró por debajo del 70% de atención.



Gráfica 9. Porcentaje de atención de los 20 usuarios

En la gráfica 10, se muestra el tiempo en el que los participantes miraron hacia la pantalla, así como el tiempo mínimo establecido para interferir atención.

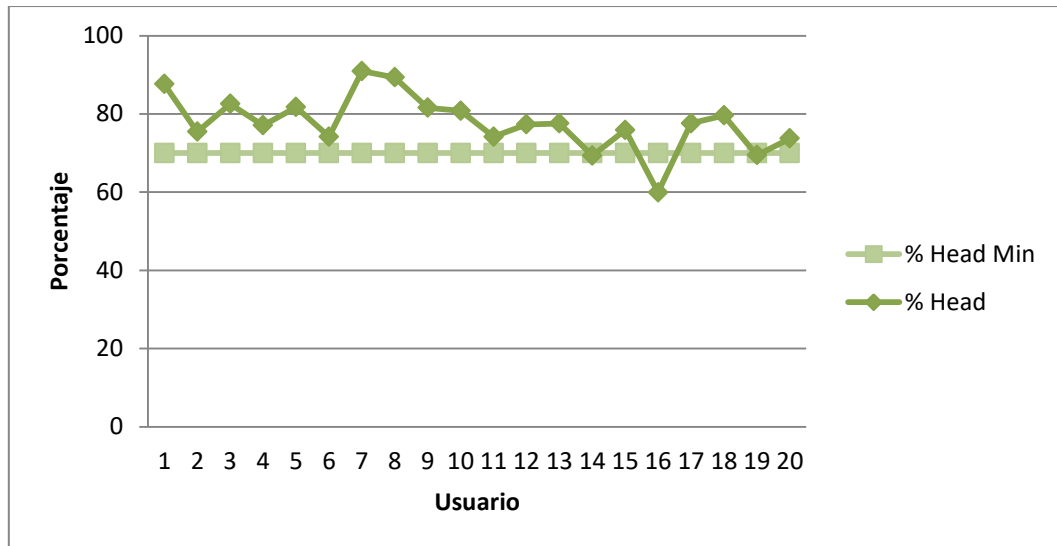
Como se puede observar el tiempo de atención disminuyó en algunos participantes ya que se encontraban demasiado cerca de la pantalla, por lo cual su rostro no era detectado en algunos frames, tal es el caso del usuario 7 11, 18 y 19.



Gráfica 10. Tiempo de atención durante la prueba

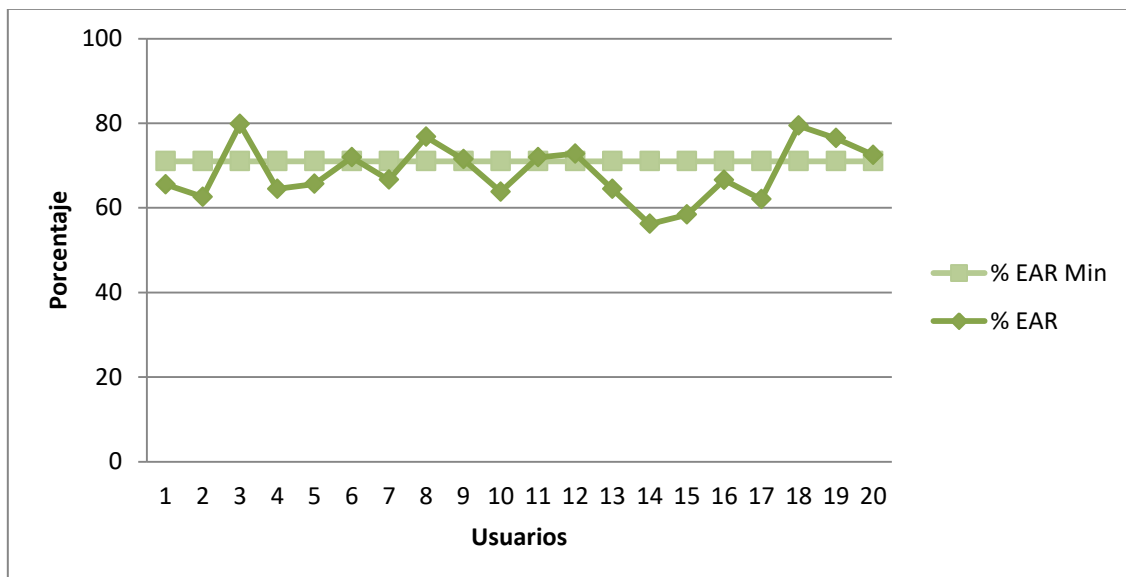
En la gráfica 11, se muestran los porcentajes promedio de los movimientos de la cabeza obtenidos por los participantes durante la actividad realizada. Pudiendo observar que en el 95% de los usuarios los movimientos de la cabeza se encuentran entre los límites establecidos.

El usuario 16 obtuvo un bajo nivel de atención respecto a la posición de su cabeza, ya que este se encontraba con la pantalla ligeramente inclinada hacia arriba, lo que dio como resultado un porcentaje bajo de atención.



Gráfica 11. Porcentaje de movimientos de la cabeza de la tarea 1

En la gráfica 12, se presenta el promedio de EAR de cada usuario con respecto a la apertura ocular mínima, al igual que en las demás tareas, este valor varia demasiado de persona a persona, encontrando que en este caso el 55% de los usuarios mantienen una apertura ocular por debajo del umbral.



Gráfica 12. Porcentaje de EAR de los usuarios

En la Tabla 6 se muestra el número de parpadeos realizados por los usuarios, el 90% se mantienen dentro de los 15 y 40 parpadeos por los 4 minutos de la prueba. Es decir el promedio los usuarios realizaron 7 u 8 parpadeos por minuto.

Usuario	# de Parpadeos
1	14
2	34
3	23
4	66
5	29
6	25
7	39
8	20
9	22
10	64
11	22
12	19
13	40
14	20
15	28
16	23
17	44
18	19
19	20
20	19

Tabla 6. Parpadeos realizados por los usuarios durante la prueba

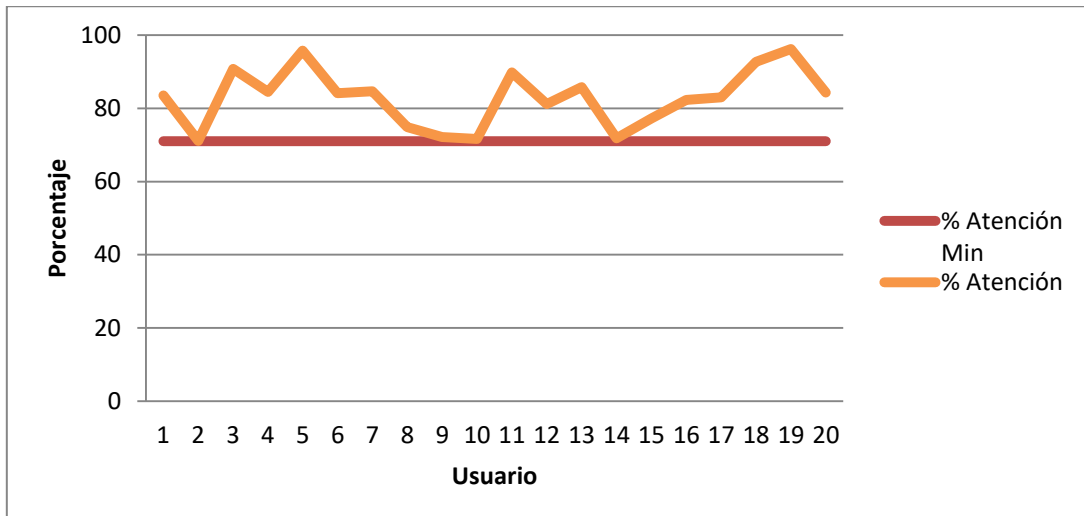
5.2.2 Tarea 2

En la tabla 7, se muestra el número de usuario, seguido por LA que muestra el porcentaje de atención de cada usuario, así como T, que expresa el tiempo en el que el usuario se mantuvo atento, el cual se presenta en minutos seguido por segundos.

Usuario	Tarea 2		Usuario	Tarea 2	
	% Atención	Tiempo		% Atención	Tiempo
1	83.50%	3 min 90s	11	89.87%	3 min 09s
2	71.14%	3 min 54s	12	81.16%	3 min 14s
3	90.84%	3 min 10s	13	85.78%	3 min 08s
4	84.52%	2 min 58s	14	71.80%	3 min 31s
5	95.82%	3 min 19s	15	77.26%	2 min 28s
6	84.19%	3 min 08s	16	82.31%	3 min 11s
7	84.69%	3 min 10s	17	82.98%	3 min 21s
8	74.89%	3 min 40s	18	92.71%	3 min 45s
9	72.20%	3 min 43s	19	96.25%	3 min 10s
10	71.66%	3 min 41s	20	84.33%	3 min 12s

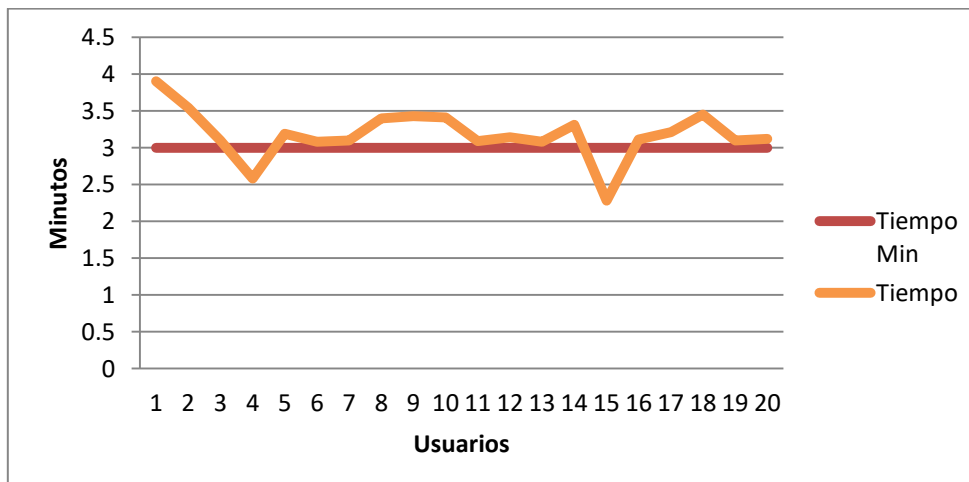
Tabla 7. Resultados mostrados por el sistema para tarea 2

En la gráfica 13, se muestran el nivel de atención que los usuarios prestaron a la actividad 2, en la que se les pidió a los usuarios utilizar el sistema mientras se encontraban leyendo algún artículo o documento académico o laboral. Como se puede observar, los niveles de atención del 100% de los usuarios se mantienen por encima del mínimo establecido.



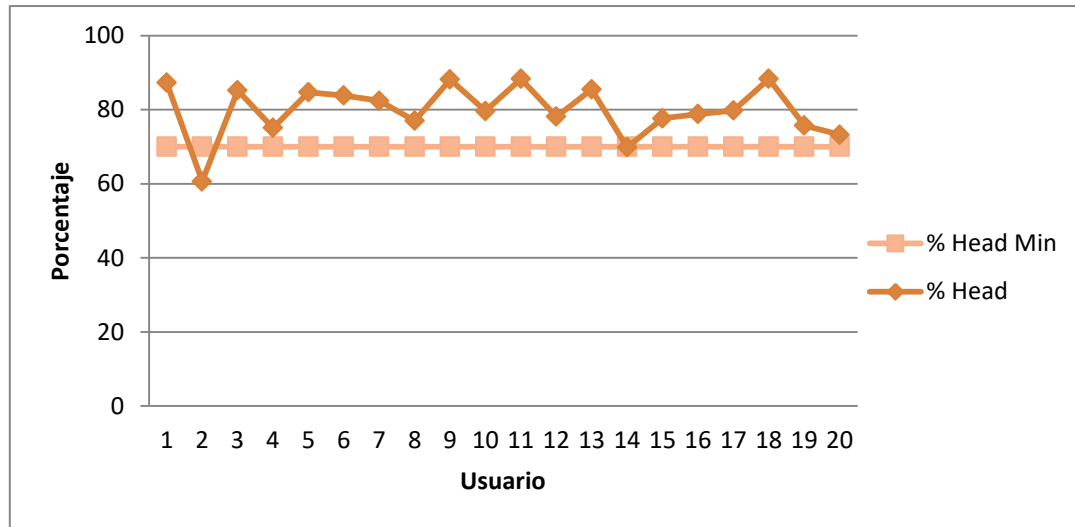
Gráfica 13. Nivel de atención de los usuarios a la prueba 2

En la gráfica 14, se muestra el tiempo en el que los usuarios se mantuvieron enfocados en la pantalla. En el cual solo el 10% de las personas estuvieron atentos menos de 3 minutos.



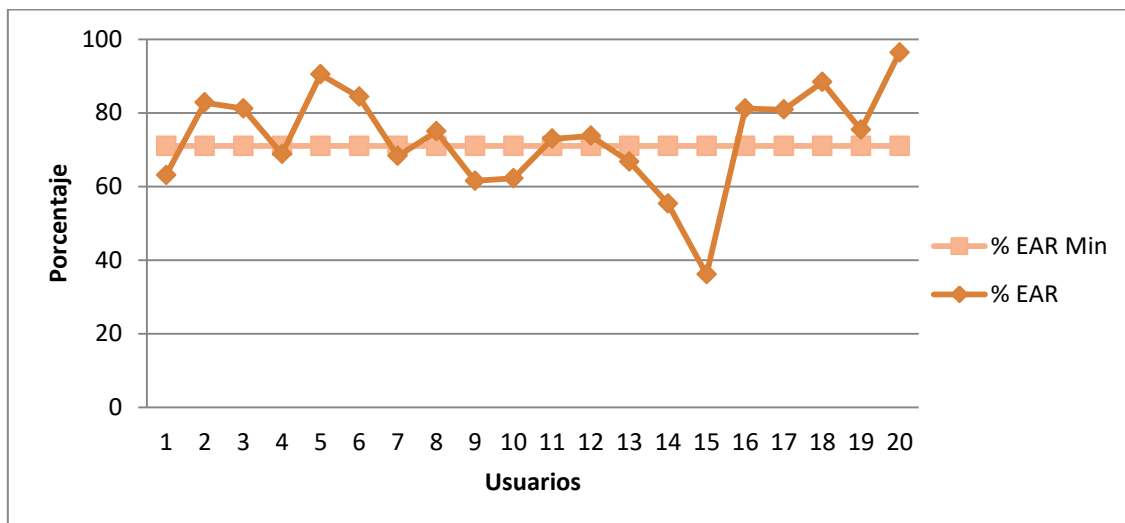
Gráfica 14. Tiempo de atención de los usuarios

En la gráfica 15 se muestran el porcentaje de los movimientos de la cabeza realizados por cada usuario, se puede observar que el promedio de los 3 ángulos de movimiento son mayormente por encima del mínimo. Solo el 5% es decir, el usuario 2 marco un porcentaje bajo debido a su incorrecto posicionamiento de la cámara.



Gráfica 15. Movimientos de la cabeza de los usuarios

El nivel de apertura ocular de los usuarios se muestra en la gráfica 16, pudiendo observar que como se ha mencionado, este valor es el que más varía. Teniendo un 40% de usuarios por debajo del umbral.



Gráfica 16. EAR de los 20 usuarios para la tarea 2

Con respecto a los parpadeos realizados en esta tarea, podemos observar en la tabla 8, los parpadeos se mantienen dentro del rango establecido, a comparación de las demás tareas puestas a prueba, en esta es en la que se registra un menor número de parpadeos, debido a que de acuerdo a la bibliografía consultada, nos concentramos tanto en lo que estamos viendo, que olvidamos realizar parpadeos.

Usuario	# de Parpadeos
1	19
2	13
3	10
4	41
5	4
6	9
7	43
8	12
9	29
10	31
11	25
12	15
13	20
14	52
15	24
16	6
17	7
18	6
19	10
20	2

Tabla 8. Numero de parpadeos de los usuarios registrados en la tarea 2

5.2.3 Tarea 3

Para esta tarea se les solicito a los participantes utilizar el programa cuando estuvieran viendo algún video de su preferencia, mirando redes sociales o interactuando con algún video juego en la computadora.

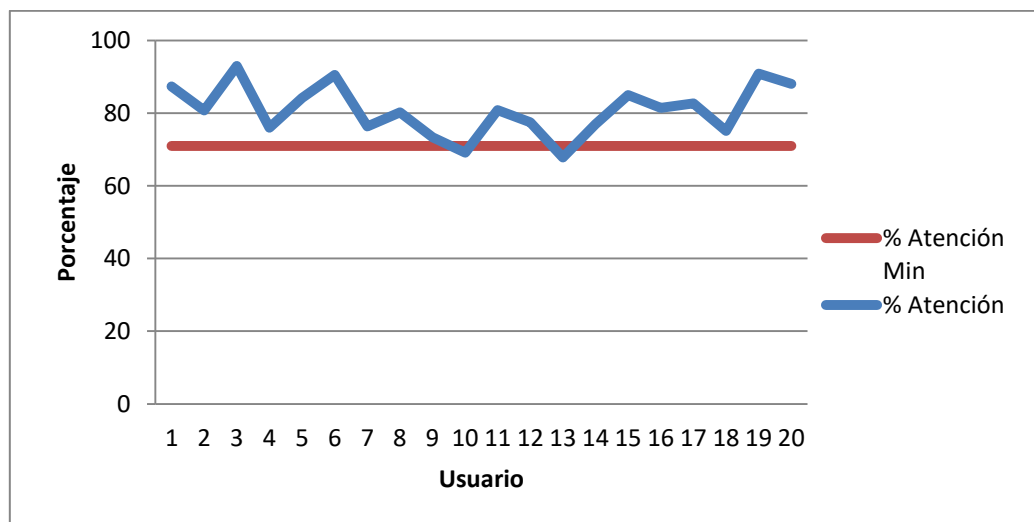
En la tabla 9, se muestra el porcentaje de atención de cada usuario así como el tiempo en el que centraron su atención a lo que se emitía de su pantalla.

En esta prueba los usuarios muestran los niveles más altos de atención con respecto a las demás tareas analizadas. Así como un tiempo mayor de atención.

Usuario	Tarea 3		Usuario	Tarea 3	
	% Atención	Tiempo		% Atención	Tiempo
1	87.29%	3 min 33s	11	80.85%	3 min 27
2	80.70%	3 min 41s	12	77.43%	3 min 25
3	92.98%	3 min 25s	13	67.82%	3 min 08
4	75.94%	3 min 05s	14	76.91%	3 min 32
5	84.19%	3 min 26s	15	84.96%	3 min 17
6	90.54%	3 min 30s	16	81.42%	3 min 48
7	76.31%	3 min 06s	17	82.67%	3 min 42
8	80.22%	3 min 37s	18	75.10%	3 min 35
9	73.30%	3 min 47s	19	90.84%	2 min 46
10	69.08%	2 min 14s	20	87.99%	2 min 52

Tabla 8. Nivel de atención de los usuarios en la tarea 3

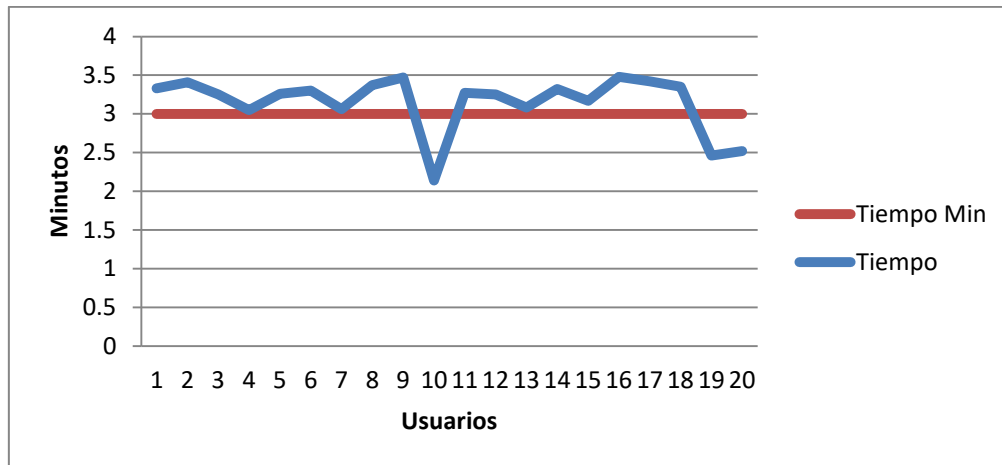
Como se puede observar en la gráfica 17, el 90% de usuarios presta atención a lo que estaban visualizando en pantalla, se refiere que solo 2 usuarios están por debajo del 70% de atención en esta actividad.



Gráfica 17. Nivel de atención 3ra tarea

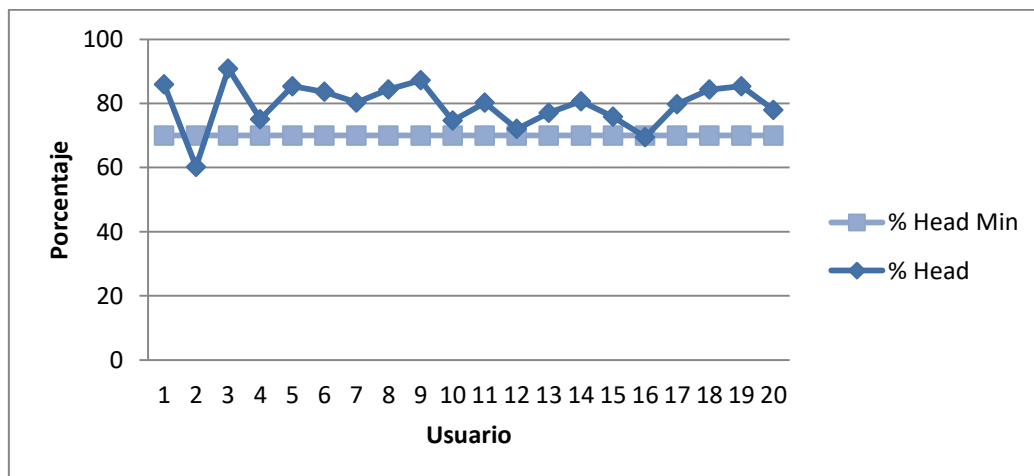
En la gráfica 18, se muestra el tiempo en el que los usuarios se mantuvieron mirando hacia la pantalla contra el tiempo mínimo establecido para interferir atención, el cual en específico de las pruebas es de 3 minutos, ya que este es el

25% del tiempo total. Cabe destacar que los usuarios con menor tiempo de atención son los usuarios que estaban mal posicionados, a los cuales no fue posible detectar su rostro en algunos frames por ello su tiempo de atención bajo.



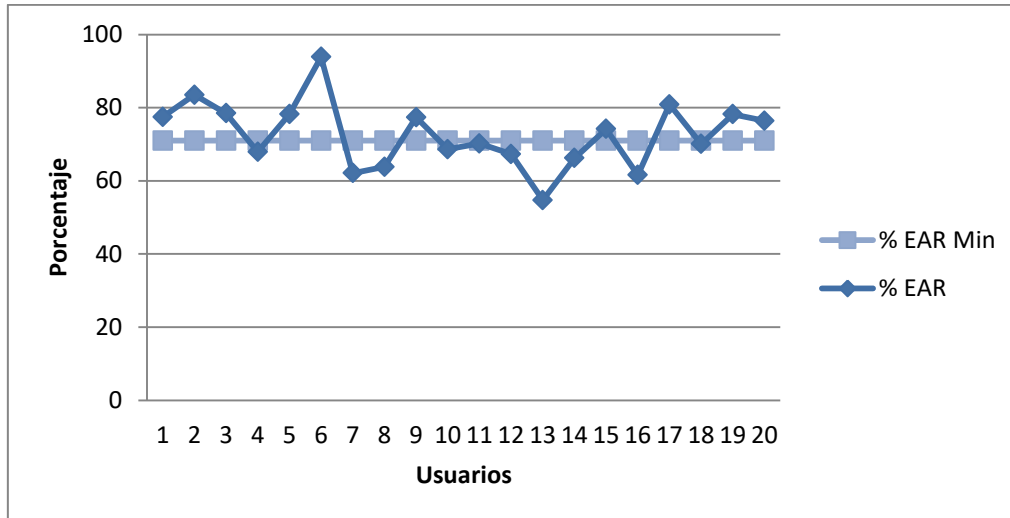
Gráfica 18. Tiempo de atención de los participantes en tarea 3

En la gráfica 19, se muestran los porcentajes de variaciones de los 3 movimientos de la cabeza obtenidas por los participantes durante dicha tarea. Pudiendo determinar que sus movimientos de la cabeza se encuentran por encima del margen establecido, encontrando solo en 2 participantes un nivel inferior al umbral, mayormente en el usuario 2 con un nivel bajo de movimientos de la cabeza, esto debido a que su rostro se encontraba en un mal ángulo, ocasionando que sus movimientos marcaran un bajo nivel de atención.



Gráfica 19. Porcentaje de movimientos de la cabeza

En la gráfica 20, se presenta el porcentaje de la relación de aspecto ocular en el cual podemos observar y como ya se ha mencionado que este resultado es el que más varia, pues hay una gran diversidad de tamaños de ojos.



Gráfica 20. Porcentaje EAR prueba 3

En la Tabla 9, se muestra el número de parpadeos realizados por los usuarios, en la cual el 90% de usuarios se mantiene en un parpadeo normal, es decir entre 5 y 10 parpadeos por minuto, que para el tiempo especificado para la prueba, son entre 15 y 40 parpadeos.

Usuario	# de Parpadeos
1	31
2	13
3	12
4	10
5	18
6	17
7	13
8	30
9	23

10	40
11	24
12	24
13	56
14	22
15	14
16	39
17	16
18	46
19	20
20	10

Tabla 9. Numero de parpadeos realizados en la 3ra tarea

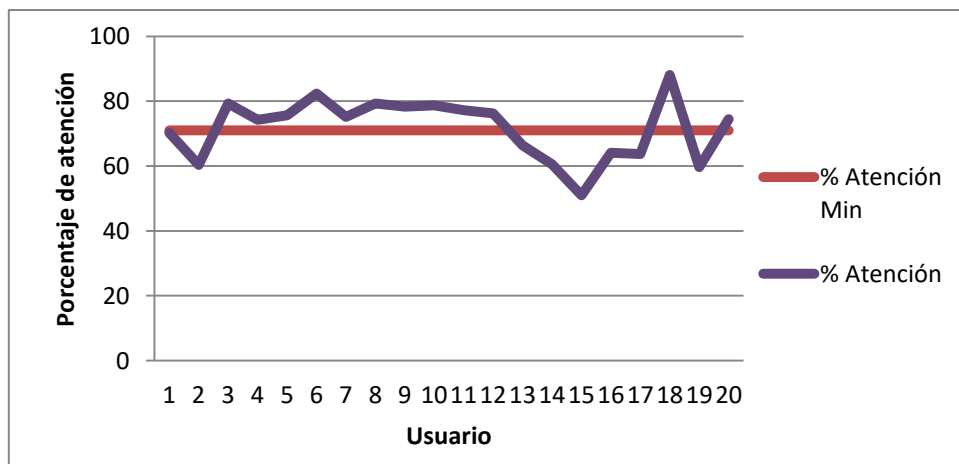
5.2.4 Tarea 4

Para esta prueba se les pidió a los participantes, utilizar el sistema cuando estuvieran escribiendo algún artículo o documento tanto académico como laboral o bien trabajando mientras escribían en su computadora.

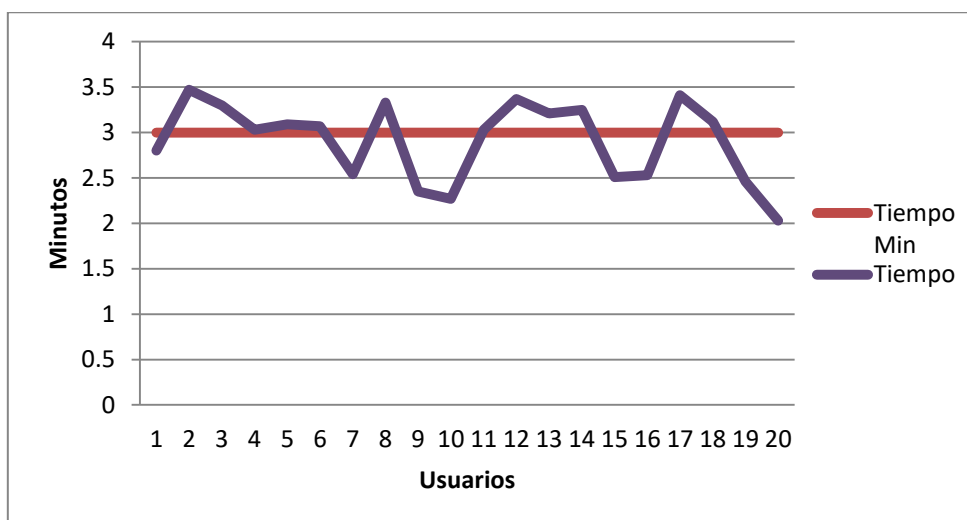
Usuario	Tarea 4		Usuario	Tarea 4	
	% Atención	Tiempo		% Atención	Tiempo
1	70.23%	2 min 08s	11	77.20%	3 min 03s
2	60.35%	3 min 47s	12	76.21%	3 min 37s
3	79.33%	3 min 30s	13	66.37%	3 min 21s
4	74.25%	3 min 03s	14	60.55%	3 min 25s
5	75.63%	3 min 09s	15	50.95%	2 min 51s
6	82.37%	3 min 07s	16	64.10%	2 min 53s
7	75.13%	2 min 54s	17	63.65%	3 min 41s
8	79.27%	3 min 33s	18	88.07%	3 min 12s
9	78.28%	2 min 35s	19	59.61%	2 min 46s
10	78.66%	2 min 27s	20	74.46%	2 min 03s

Tabla 10. Resultados obtenidos del sistema para la tarea 4

Como se puede observar en la tabla 10, los porcentajes de atención en comparación de las otras 3 tareas, son los más bajos debido a que la apertura ocular es menor y los parpadeos registrados se incrementan.

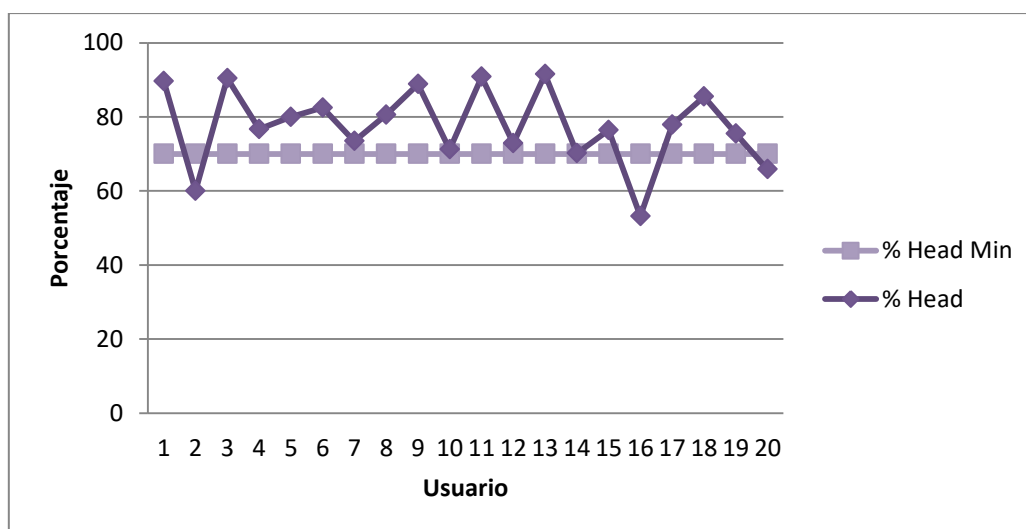


Gráfica 21. Nivel de atención de la tarea 4



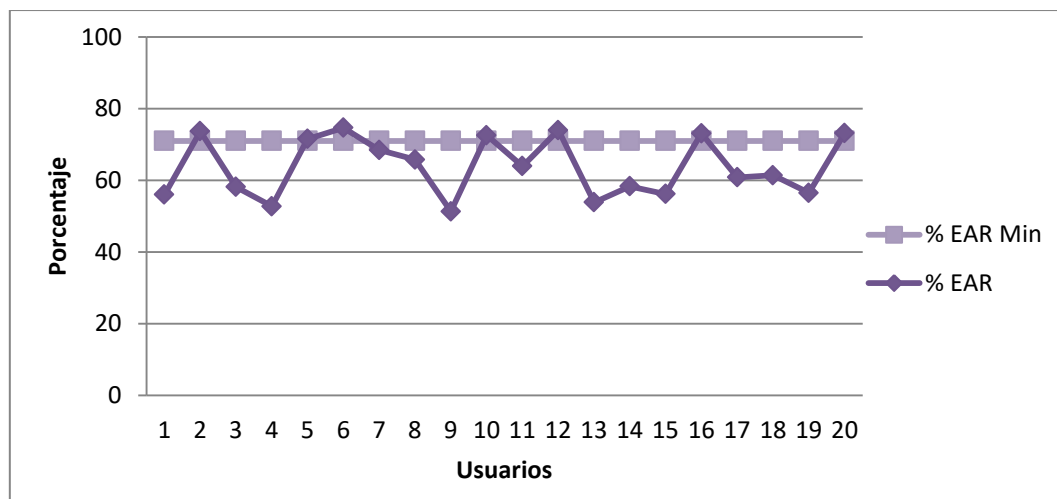
Gráfica 22. Tiempo de atención de los participantes para la 4ta prueba

En la gráfica 22, se observa el tiempo de atención de los usuarios en comparación al tiempo mínimo para inferir que se presta atención.



Gráfica 23. Porcentaje de movimiento de la cabeza

Como se puede observar en la gráfica anterior, solo el 15% de los usuarios se mantuvieron por debajo de los movimientos establecidos.



Gráfica 24. Porcentaje de EAR de la 4ta tarea

En la gráfica 24, se puede observar que el porcentaje de apertura ocular se encuentra mayormente por debajo del valor establecido, esto debido a que al realizar la actividad de escritura se baja la mirada hacia el teclado, ocasionando que el EAR baje demasiado en la mayoría de los participantes.

En la tabla 11, se muestran los parpadeos realizados durante esta prueba, resaltando que como la apertura ocular baja al mirar el teclado y se mantiene así mientras se escribe, se registran más parpadeos, al igual que por la acción de mirar el teclado y volver a mirar la pantalla. Se registran el 45% de los usuarios con parpadeos por encima de los establecidos como normales.

Usuario	# de Parpadeos
1	17
2	49
3	55
4	31
5	66
6	70
7	40
8	13
9	38

10	23
11	27
12	32
13	78
14	48
15	29
16	42
17	56
18	10
19	58
20	17

Tabla 11. Numero de parpadeos realizados por los usuarios en la 4ta tarea

5.2.5 Análisis de los resultados

De acuerdo a los resultados tanto cualitativos (descritos por los usuarios) y los cuantitativos (obtenidos por el sistema), el sistema obtuvo un nivel de atención acorde a la retroalimentación dada por el usuario, a los cuales se les solicitó dar un breve informe sobre sí mismo mientras realizaban la prueba, indicando de manera sincera si habían estado prestando atención o no.

Es decir, el sistema predice de manera correcta si el usuario pone atención a lo presentado en la pantalla, más específicamente de acuerdo a sus movimientos tanto de la cabeza como de sus ojos puede indicar un porcentaje de atención que se presta a las actividades realizadas en la computadora.

Conclusiones

A lo largo del trabajo realizado, se investigaron los parámetros “normales” en los cuales se mantiene una persona al trabajar en un equipo de cómputo. El sistema detecta de manera eficiente el nivel de atención con respecto a los movimientos realizados, así como el tiempo que los usuarios estuvieron atentos a la pantalla.

Como se mostró el nivel de atención así como el tiempo varía dependiendo la actividad, donde de acuerdo a las pruebas realizadas la actividad a la que más le prestaron atención los participantes fue a la tarea 2, teniendo un promedio del 82.90% de atención, yendo acorde a que la prueba de lectura requiere un mayor concentración debido a que se busca comprender lo que se está leyendo, la siguiente prueba con mayor número de atención es la prueba 3 con un 80.83% ya que se presta más atención cuando se está realizando una actividad que nosotros elegimos, la actividad 1 en la que se encontraban en algún enlace de video reflejo un 80.16% muy cerca de la prueba anterior, con un 71.73% se encontró la prueba de escritura que como se ha comentado es la que implica mayor movilidad de la cabeza y menor apertura ocular para realizar dicha tarea.

Como se pudo observar, lo que más variaba es la EAR, debido a que cada usuario tiene un tamaño y tipo de ojo diferente, y esto si influye en el resultado ya que se establece un margen fijo para todos los usuarios. Y esto ocasiono que en la mayoría de los usuarios, su nivel de atención bajara.

Del mismo modo se concluye que el proyecto es funcional cuando lo que se requiere es saber si miran la pantalla, y para esta tarea en específico el sistema es viable, ya que detecta de manera eficiente el rostro de los usuarios aun en condiciones donde existe poca luz, así como detecta los ojos de las personas aun cuando estas usan lentes.

Trabajo futuro

Para hacer el sistema mayormente eficiente, se podría realizar algunos ajustes al programa principalmente en tema de EAR, puesto que, como fue mencionado la apertura ocular varia demasiado debido al tamaño de los ojos de cada usuario, para lo cual se podría realizar una medición previa al uso del sistema para determinar la atención.

Se tendría que grabar un par de segundos al usuario para poder obtener la apertura máxima promedio del usuario, lo que nos daría un umbral personalizado para poder determinar de manera exacta cuando se produce un parpadeo o bien cuando la persona de acuerdo a su apertura ocular refleje cansancio, sueño o algún otro factor relacionado con una baja apertura ocular. Este valor podría ser generado una sola vez y guardado para así poder identificar cuando este usuario vuelva a medir su porcentaje de atención. Esto nos daría una EAR más acertada ya que se podrá detectar de manera correcta la apertura ocular de cada usuario.

Del mismo modo, se le podría integrar un reconocimiento facial, de este modo no solo se medirá el nivel de atención, si no, que también se detectara que usuario lo está usando, esto será de utilidad a la hora de realizar exámenes en línea, ya que se tendrá la certeza de que la persona que lo solicito sea la misma que lo presenta, así como verificar que todo el tiempo su atención este hacia lo que está en pantalla.

Referencias

- [1] Li, S. Z., & Jain, A. K. (2011). Handbook of Face Recognition. Springer Publishing.
- [2] É. K. Shimomoto, A. Kimura and R. Belém, (2015). A faster face detection method combining Bayesian and Haar Cascade Classifiers. Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON), Santiago, pp. 7-12, <http://doi.org/10.1109/Chilecon.2015.7400344>.
- [3] Siddiqi, M.H., Ali, R., Khan, A.M. et al. (2015) Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection. Multimedia Systems 21, 541–555. <https://doi.org/10.1007/s00530-014-0400-2>
- [4] D. Marčetić, T. Hrkać and S. Ribarić, "Two-stage cascade model for unconstrained face detection (2016), First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE), Aalborg, 2016, pp. 1-4, doi: 10.1109/SPLIM.2016.7528404.
- [5] Mutneja, V., Singh, S (2019). Modified Viola–Jones algorithm with GPU accelerated training and parallelized skin color filtering-based face detection. J Real-Time Image Proc 16, 1573–1593.
- [6] Lu, Y.; Zheng, WL; Li, B.; Lu, (2015) Combining Eye Movements and EEG to Enhance Emotion Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.
- [7] M. N. Chaudhari, M. Deshmukh, G. Ramrakhiani and R. Parvatikar,(2018). Face Detection Using Viola Jones Algorithm and Neural Networks, Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697768.
- [8] Narayanan Veliyath, Pradipta De, Andrew A. Allen, Charles B. Hodges, and Aniruddha Mitra. (2019). Modeling Students' Attention in the Classroom using Eyetrackers. In Proceedings of the 2019 ACM Southeast Conference (ACM

- SE '19). Association for Computing Machinery, 2–9.
DOI:<https://doi.org/10.1145/3299815.3314424>
- [9] Menchon, M. (2011). Implementación de Seguimiento de Rostros en Rehabilitación Cognitiva.
- [10] R. Oyini Mbouna, S. G. Kong and M. -G. Chun, (2013). Visual Analysis of Eye State and Head Pose for Driver Alertness Monitoring. In IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 3, pp. 1462-1469, Sept. 2013, Doi: 10.1109/TITS.2013.2262098.
- [11] N. Ruiz, E. Chong and J. M. Reh. (2018). Fine-Grained Head Pose Estimation Without Keypoints. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Doi: 10.1109/CVPRW.2018.00281.
- [12] M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan. (2003). Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction., Conference on Computer Vision and Pattern Recognition Workshop, pp. 53-53, doi: 10.1109/CVPRW.2003.10057.
- [13] S. Duffner and C. Garcia, (2016). Visual Focus of Attention Estimation With Unsupervised Incremental Learning," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, pp. 2264-2272, doi: 10.1109/TCSVT.2015.2501920.
- [14] X. Zhang, C. Wu, P. Fournier-Viger, L. Van and Y. Tseng, (2017). Analyzing students' attention in class using wearable devices. IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1-9, doi: 10.1109/WoWMoM.2017.7974306.
- [15] N. Krishnnan, S. Ahmed, T. Ganta and G. Jeyakumar, (2020). A Video Analytics Based Solution for Detecting the Attention Level of the Students in Class Rooms, 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 498-501, doi: 10.1109/Confluence47617.2020.9057967.

- [16] Kaspersky. (2021). *Reconocimiento facial: definición y explicación*. latam.kaspersky.com.<https://latam.kaspersky.com/resource-center/definitions/what-is-facial-recognition>
- [17] Klette, R. (2014). *Concise computer vision: An introduction into theory and algorithms*. Springer, 375-413.
- [18] Viola, P., & Jones, M. (2001). Robust real-time face detection. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV*.
- [19] Ming-Hsuan Yang, D. J. Kriegman and N. Ahuja, "Detecting faces in images: a survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, Jan. 2002, doi: 10.1109/34.982883.
- [20] *i-bug - resources - Facial point annotations*.
<https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/>.
- [21] M. Uříčář, F. Vojtěch y V. Hlavac, (2012) Detector of facial landmarks learned by the structured output svm, *VISAPP'12: 7th International Conference on Computer Vision Theory and Applications*.
- [22] V. Vapnik, (1995) *The Nature of Statistical Learning Theory*.
- [23] Samaria, F., & Young, S.J. (1994). HMM-based architecture for face identification. *Image Vis. Comput.*
- [24] T. F. Cootes, C. J. Taylor, D. H. Cooper y J. Graham (1995), Active shape modelstheir training and application, *Computer vision and image understanding*, vol. 61.
- [25] Antunez, M.A (2009). Algoritmos de detección de objetos para la detección y seguimiento de ojos.
- [26] V. F. Ferrario, C. Sforza, G. Serrao, G. Grassi y E. Mossi (2002). Active rang of motion of the head and cervical spine: A three-dimensional investigation in healthy young adults, *Journal of orthopaedic research*, vol. 20.
- [27] H. A. Rowley, S. Baluja y T. Kanade (1998), Rotation invariant neural networkbased face detection.
- [28] T. F. Cootes, G. J. Edwards y C. J. Taylor (2001), Active appearance models", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, n.o 6.

- [29] T. F. Cootes, C. J. Taylor, D. H. Cooper y J. Graham (1995), Active shape modelstheir training and application, Computer vision and image understanding, vol. 61.
- [30] Rocca, F., Mancas, M., Gosselin, B. (2014). Head Pose Estimation by Perspective-n-Point Solution Based on 2D Markerless Face Tracking. In: Reidsma, D., Choi, I., Bargar, R. (eds) Intelligent Technologies for Interactive Entertainment. INTETAIN 2014. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering
- [31] Vasca, N. (2020, 14 octubre). Cuáles son las posturas correctas para trabajar con una pc o laptop. Pronectis. <https://pronectis.com/novedades/cuales-son-las-posturas-correctas-para-trabajar-con-una-pc-o-laptop/>
- [32] Crespo, M. T. D., & Dapena, C. L. (2005). *Transtornos visuales del ordenador*. 3M.
- [33] Benitez, C. J. M (2020). La enfermedad del ojo seco
- [34] CogniFit. (2016, 4 febrero). Atención. Atención - Proceso Cognitivo. <https://www.cognifit.com/es/atencion>
- [35] Psise: Servicio de Psicología Clínica del Desarrollo. Unidad de Observación y Dlagnóstico Funcional. (2021, 8 enero). Atención: definición y funcionamiento - PSISE Madrid. PSISE. <https://psisemadrid.org/atencion/>
- [36] H. (2019, 26 julio). Expresión corporal: ¿Qué dicen tus gestos sobre ti? Hotmart. <https://blog.hotmart.com/es/expresion-corporal/>
- [37] R. (2021, 4 enero). ▷Psicología de la Atención: Estrategias y Recursos Atencionales. Psicocode. <https://psicocode.com/psicologia/psicologia-atencion/>
- [38] Soukupová, T., & Cech, J. (2016). Real-Time Eye Blink Detection using Facial Landmarks.