



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA



Departamento de Eléctrica, Electrónica y Telecomunicaciones

Carrera de Ingeniería en Electrónica, Automatización y Control

Tema:

Diseño, implementación y pruebas de desempeño de un sistema bajo costo para emular las funciones del ratón de un computador a través de gestos faciales utilizando algoritmos de visión artificial y reconocimiento de voz

Autor:

Kevin Gabriel Valencia Aragón

Director:

Ing. Pablo Ramos

Sangolquí - 2022

Agenda

Justificación

Objetivos

Descripción del sistema

Pruebas de desempeño

Conclusiones

Referencias



Justificación



El uso del computador ha crecido en los últimos tiempos debido a la necesidad de estar conectados. Es una herramienta que facilita en gran parte el aprendizaje debido a toda la información a la que se puede acceder.

Personas con discapacidades motrices tienen limitaciones para utilizar estas máquinas debido a que no pueden interactuar con los periféricos que permiten su control.

La implementación de un emulador basado en reconocimiento facial y reconocimiento de voz, ayudarían a este grupo de personas a aprovechar los beneficios del uso del computador.

Es indispensable buscar alternativas que sean accesibles y faciliten el manejo del computador por parte de este grupo de personas, sin la necesidad de adquirir dispositivos externos que puedan resultar incómodos y costosos para los usuarios.



Objetivos

Objetivo general

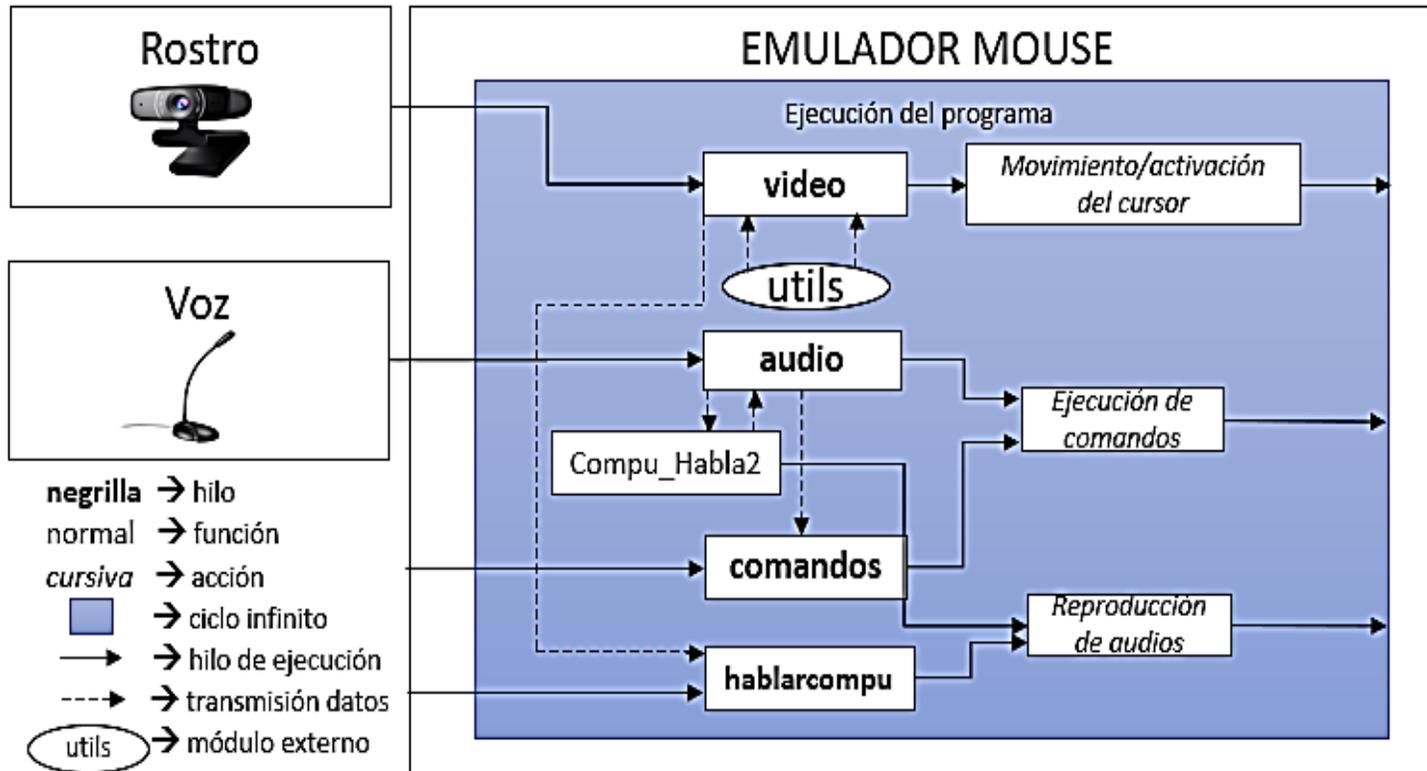
- Diseñar, implementar y validar un sistema que permita emular las funcionalidades de un ratón, utilizando algoritmos de visión artificial y de reconocimiento de voz, enfocado para personas con discapacidades motrices.

Objetivos específicos

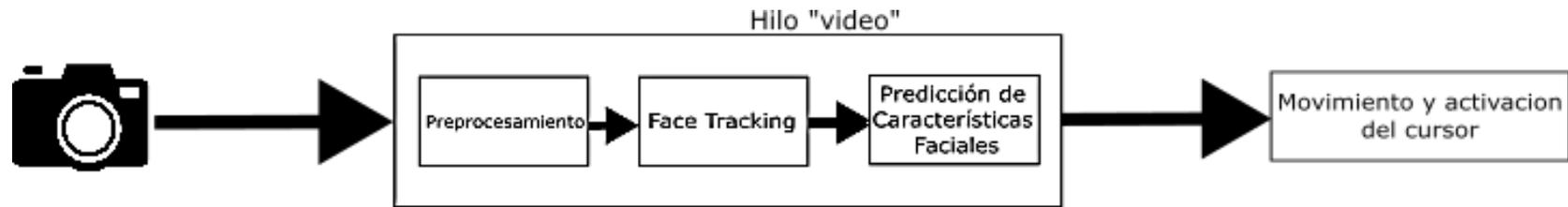
- Implementar algoritmos de visión artificial que permitan detectar diferentes zonas del rostro y usarlas para controlar el cursor.
- Utilizar algoritmos de reconocimiento de voz para facilitar el manejo del emulador.
- Desarrollar un software que combine las funcionalidades de los algoritmos de visión artificial y reconocimiento de voz para emular el funcionamiento del *mouse*.
- Realizar diferentes pruebas para medir su funcionalidad y validar su desempeño bajo distintas situaciones.



Descripción del sistema

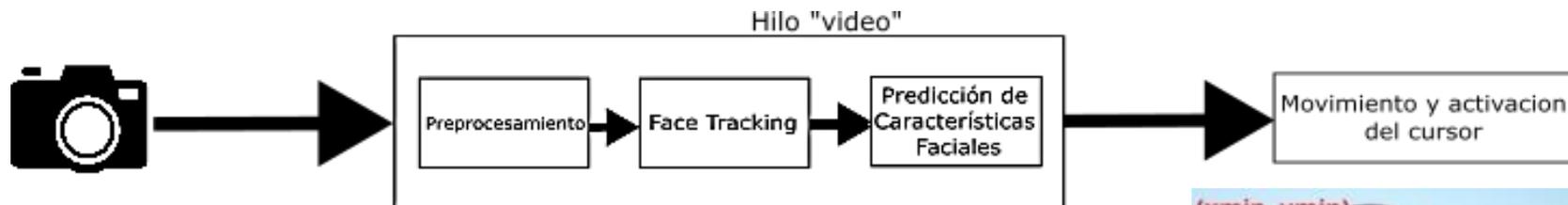


- Se presenta un sistema basado en Face tracking y Speech recognition desarrollado en python.
- Consiste en cuatro hilos que en conjunto permiten emular el funcionamiento del mouse.
- La transmisión de datos entre hilos se logra con objetos de la clase Queue los cuales permiten almacenar y distribuir datos de forma ordenada. Se trabajan con estructuras FIFO (First in First out).

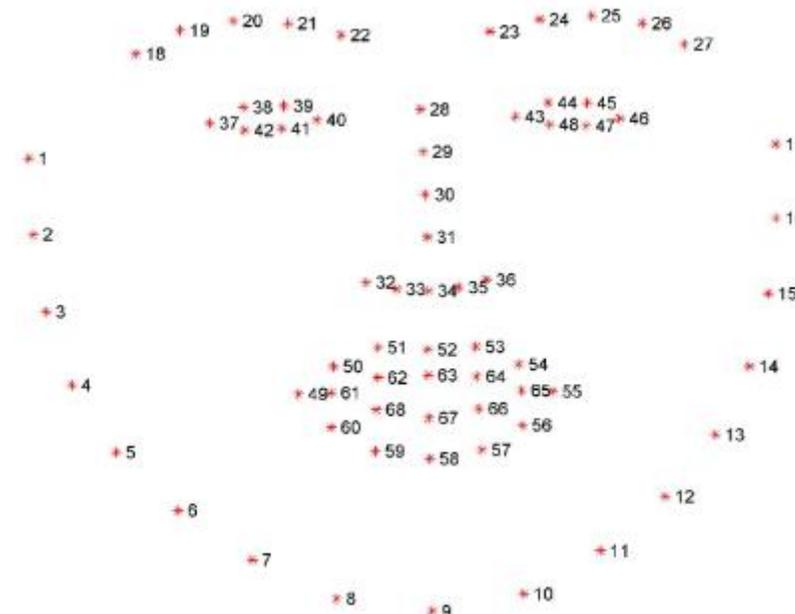
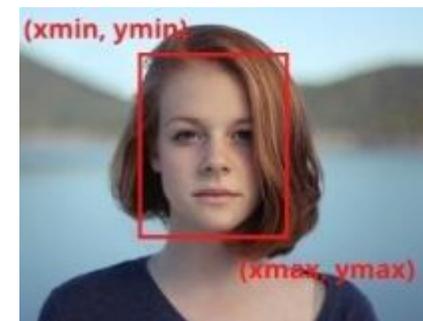


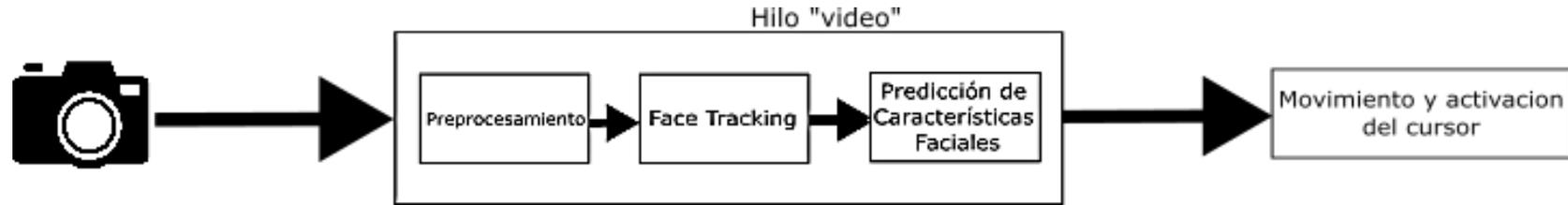
- Se adquieren las imágenes de la cámara del usuario y pasan por un bloque de preprocesamiento.
 - Se realiza un reflejo Horizontal de la imagen.
 - Se transforma a escala de grises.



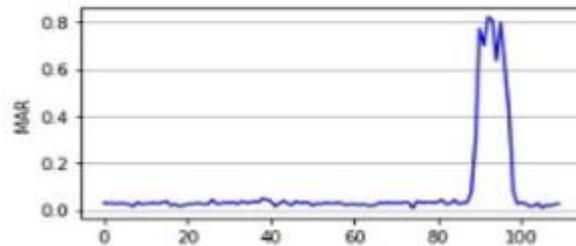
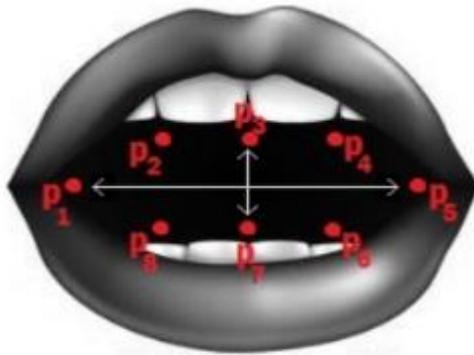
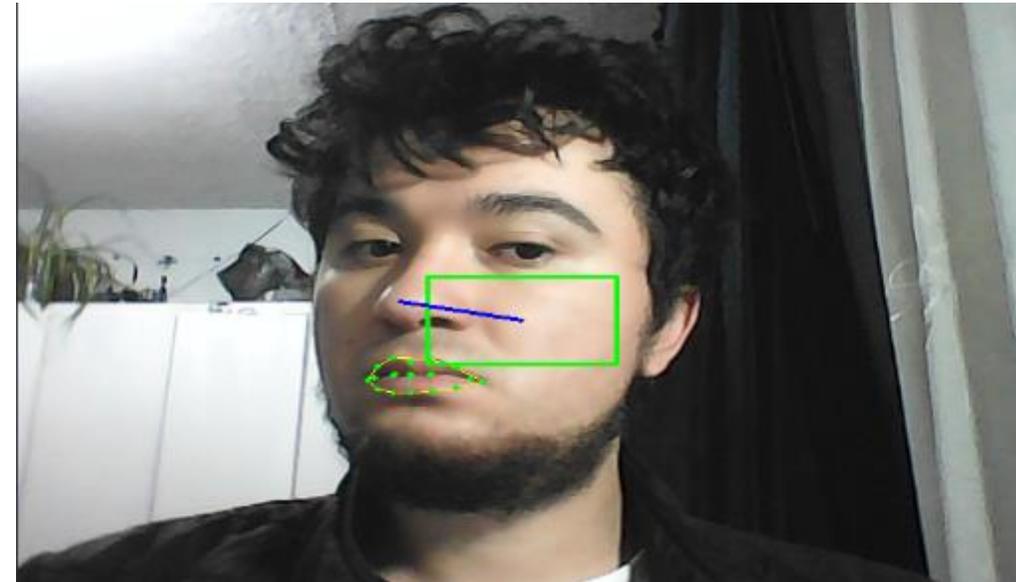


- Sobre la imagen en escala de grises se aplica el detector de rostros de la librería Dlib. Que da como resultado una matriz que contiene las coordenadas de los extremos de un rectángulo que encierra al rostro del usuario.
- Sobre el rostro detectado se aplica el predictor de características faciales de la librería Dlib el cual da resultados en tiempo real y de alta calidad.
- Para los fines del emulador se utilizan únicamente las coordenadas de la boca y del punto 34 correspondiente a la nariz.



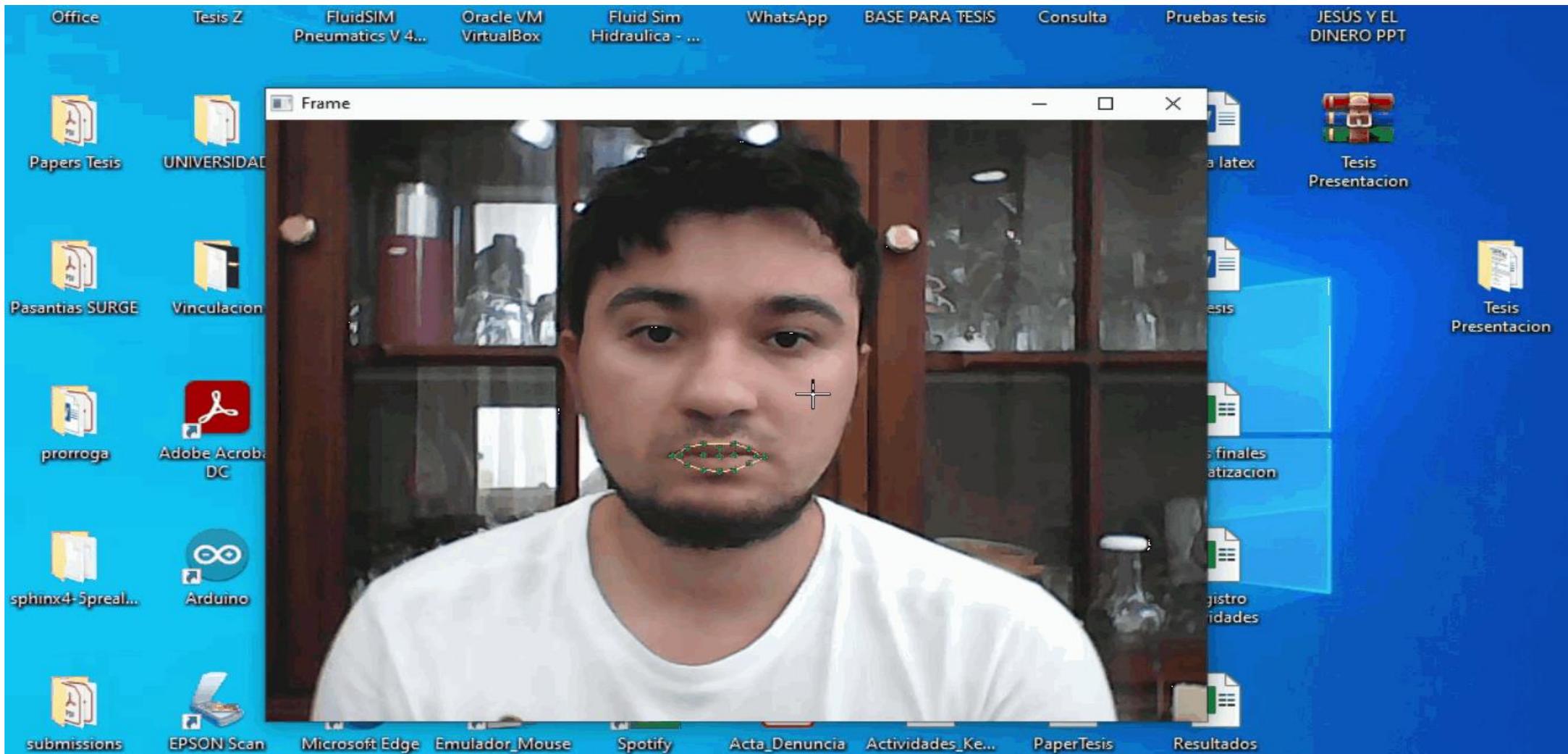


- Con las coordenadas de la boca, se implementó un sistema para activar y desactivar el movimiento del cursor.
- El movimiento del cursor se logró utilizando la coordenada del punto central de la nariz.

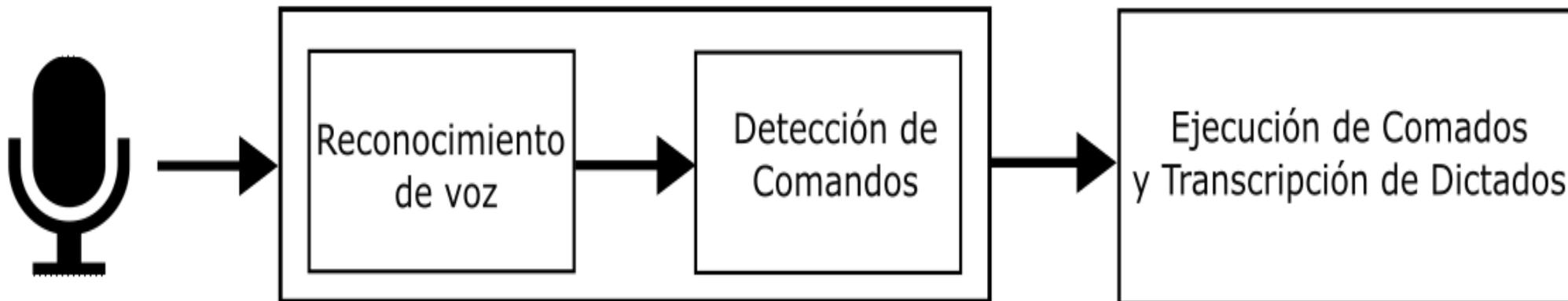


Chandra (2018)

$$MAR = \frac{\|p_2 - p_8\| + \|p_3 - p_7\| + \|p_4 - p_6\|}{2\|p_1 - p_5\|}$$

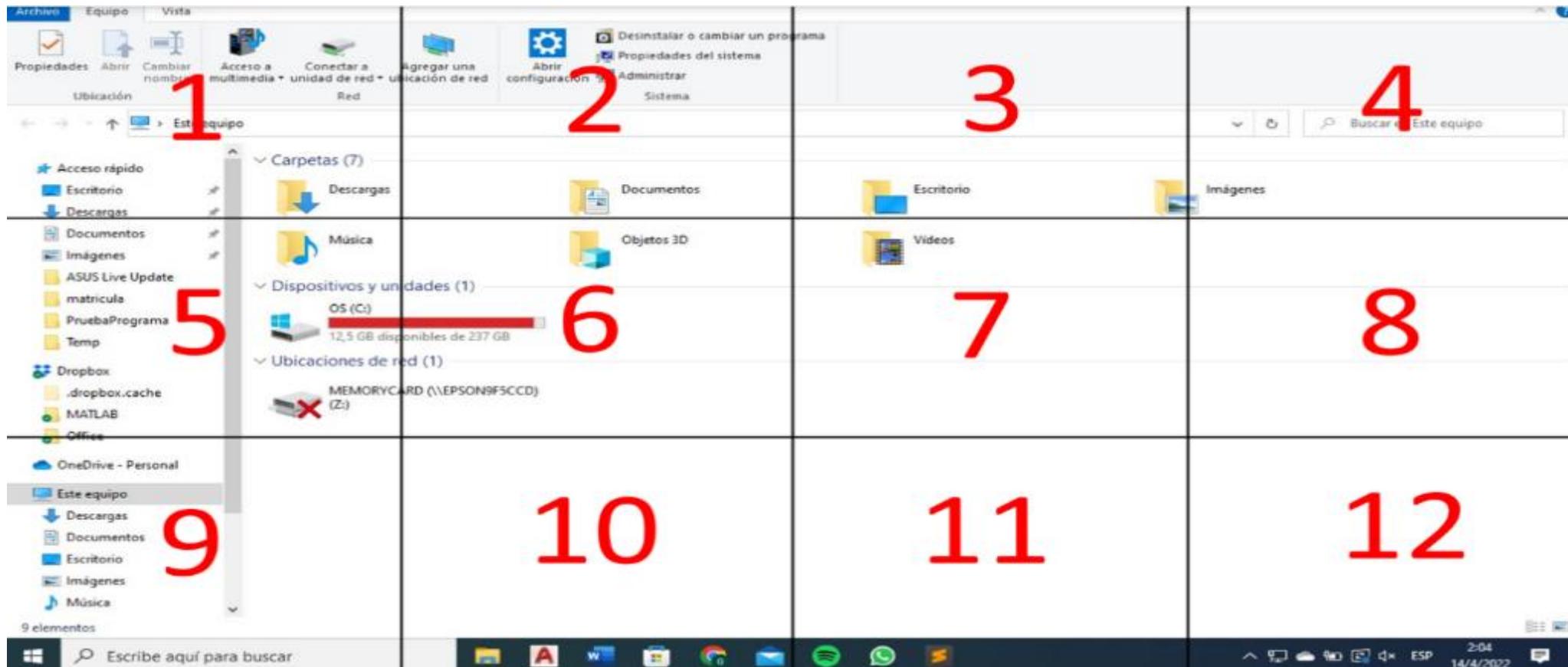


- Usando el mismo sistema para activar el movimiento del cursor, se puede resetear el punto de referencia.
- Se usó la librería Pyautogui para generar los movimientos del cursor.



- Se utiliza la librería “Vosk” para lograr el reconocimiento de la voz.
- Para lograr el reconocimiento en idioma español se utilizan modelos acústicos y de lenguaje pre entrenados a través de algoritmos de machine learning.
- Se instancia un objeto de la clase vosk que funciona como reconocedor, el mismo tiene como parámetros los modelos anteriormente descritos.
- Cuando se obtiene el texto decodificado en un string en minúsculas de lo que el usuario ha dicho, se lo compara con palabras específicas que servirán como comandos.

	Comando	Función
Segmentación De Pantalla	uno	Mueve el cursor al Cuadrante 1
	dos	Mueve el cursor al cuadrante 2
	tres	Mueve el cursor al Cuadrante 3
	cuatro	Mueve el cursor al Cuadrante 4
	cinco	Mueve el cursor al Cuadrante 5
	seis	Mueve el cursor al Cuadrante 6
	siete	Mueve el cursor al Cuadrante 7
	ocho	Mueve el cursor al Cuadrante 8
	nueve	Mueve el cursor al Cuadrante 9
	diez	Mueve el cursor al Cuadrante 10
	once	Mueve el cursor al Cuadrante 11
	doce	Mueve el cursor al Cuadrante 12
Posicionamiento	inicio	Realiza un clic en la esquina inferior izquierda, ícono de inicio (Windows/Linux).
	escritorio	Realiza un clic en la esquina inferior derecha, sobre barra de acceso rápido a escritorio en Windows y posiciona el cursor en el centro de la pantalla (Exclusivo para Windows).
	centro	Posiciona el cursor en el centro de la pantalla.
Funciones Clásicas	clic	Realiza un clic
	"doble" o "doble clic"	Realiza doble clic
	derecho	Realiza un clic derecho
	sujetar	Mantiene presionado clic izquierdo
	soltar	Deshace acción "sujetar"
Combinaciones De Teclas	cerrar	Cierra la ventana actual
	minimizar	Minimiza la ventana actual
	ventanas	Abre el menú de ventanas abiertas (Exclusivo para Windows)
	enter	Realiza la función de la tecla "ENTER"
Activación	dictado	Activa el modo dictado
	salir dictado	Desactiva el modo dictado
	apagar	Cierra el emulador



- Son implementados con la finalidad de compensar la velocidad de movimiento del cursor.
- Facilitan el movimiento a través de toda la pantalla.

- Es una función especial del hilo “audio”.
- Permite transcribir a texto todas las frases que el usuario esté diciendo.
- Los comandos de la lista general funcionan normalmente.
- Se implementan funciones exclusivas.

Comando	Función
borrar	Borra 1 caracter
borrar dos	Borra 2 caracteres
borrar tres	Borra 3 caracteres
borrar cuatro	Borra 4 caracteres
borrar cinco	Borra 5 caracteres
borrar diez	Borra 10 caracteres
borrar mucho	Borra 40 caracteres
espacio	Genera un espacio en blanco en el texto

Hilo “comandos”

- Recibe valores específicos del hilo “audio” para diferenciar comandos.
- Se encarga de ejecutar los comandos de segmentación de pantalla y de posicionamiento.
- Realiza el cálculo de las coordenadas específicas donde se debe posicionar el cursor dependiendo del comando que se ha dicho.



Hilo “hablarcompu”

- Recibe un indicador del hilo “video” para diferenciar el estado del movimiento del cursor.
- Se encarga de reproducir alertas de voz para indicar que el movimiento del cursor ha sido activado o desactivado.

La función `compu_habla2` es ejecutada desde el hilo Audio, y se encarga de reproducir alertas de voz para identificar que el modo dictado ha sido activado o desactivado.



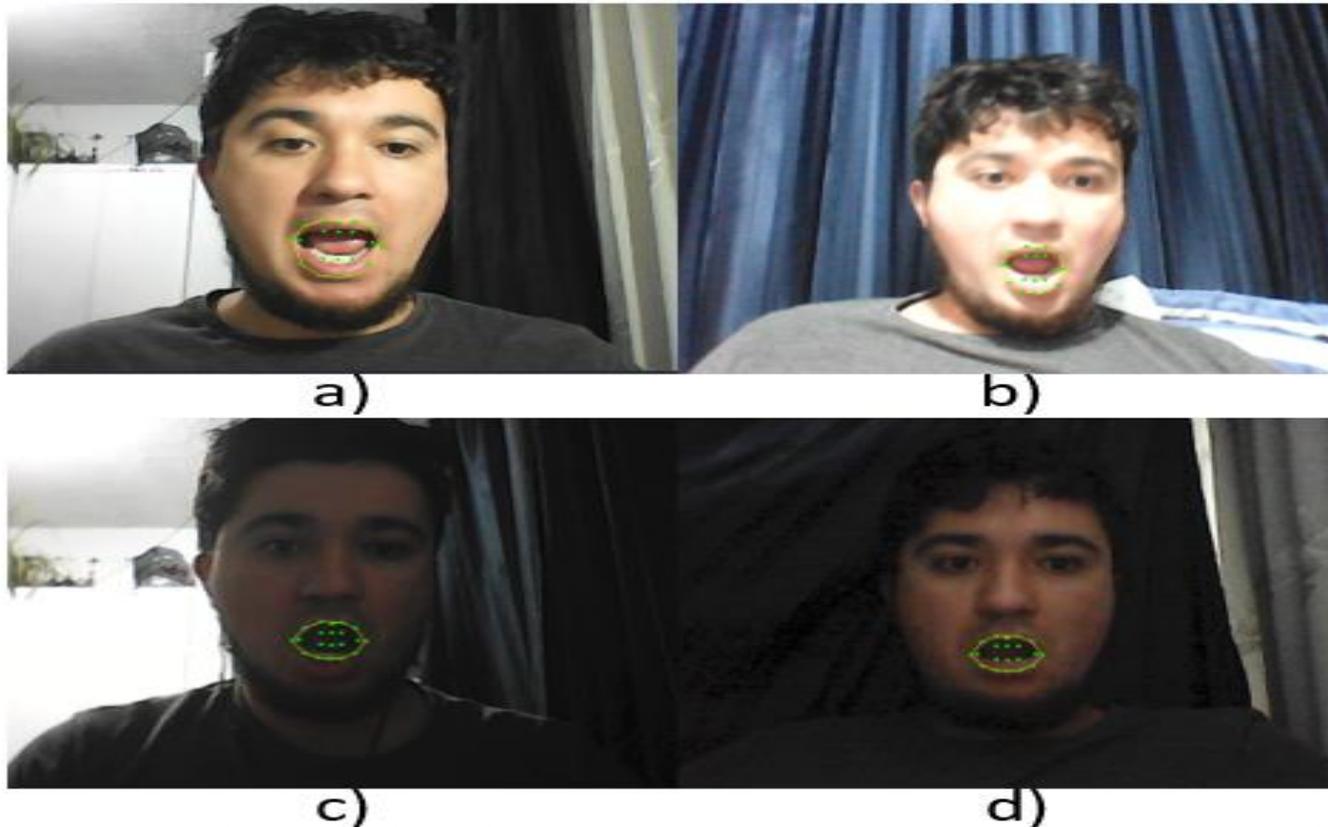


Pruebas de Desempeño



Pruebas de iluminación

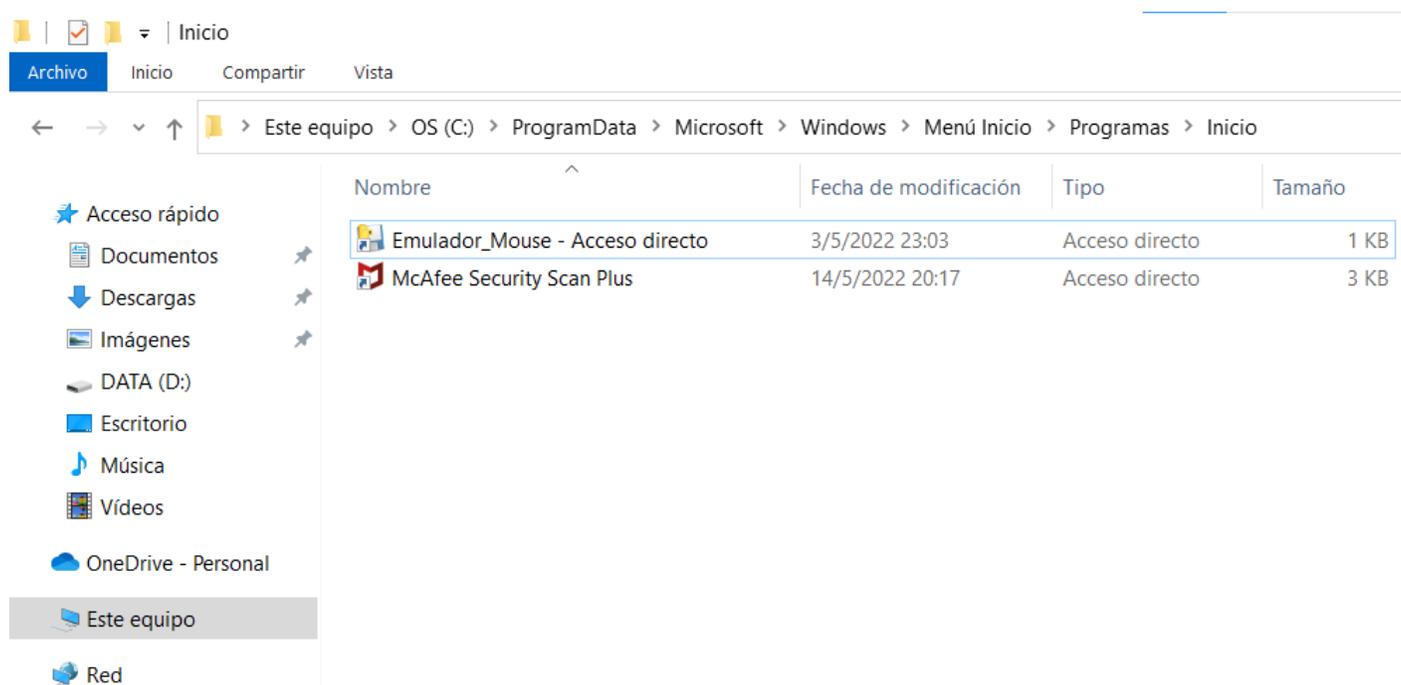
- Se probó la detección del rostro bajo distintos niveles de luminosidad.
- Se usó la webcam integrada en el computador (VGA – 640x480)



Nota. a) Imagen capturada con 469 luxes, b) Imagen capturada con 107 luxes, c) Imagen capturada a 19 luxes, d) Imagen capturada a 5 luxes

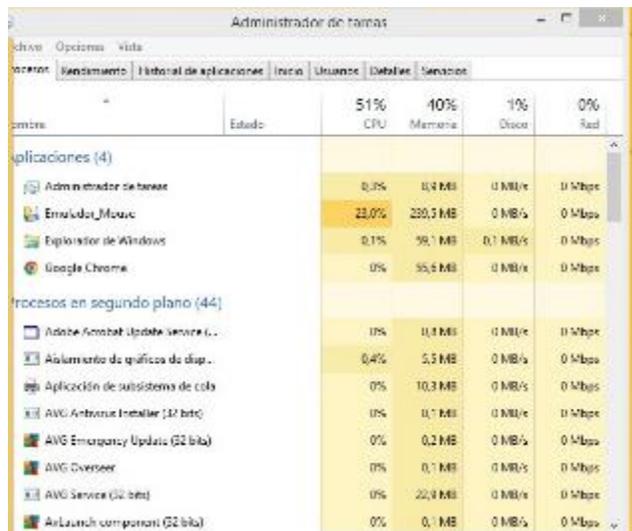
Prueba de inicio automático

- Se buscó la manera de lograr que el emulador se inicie automáticamente al momento de abrir la sesión en el computador.

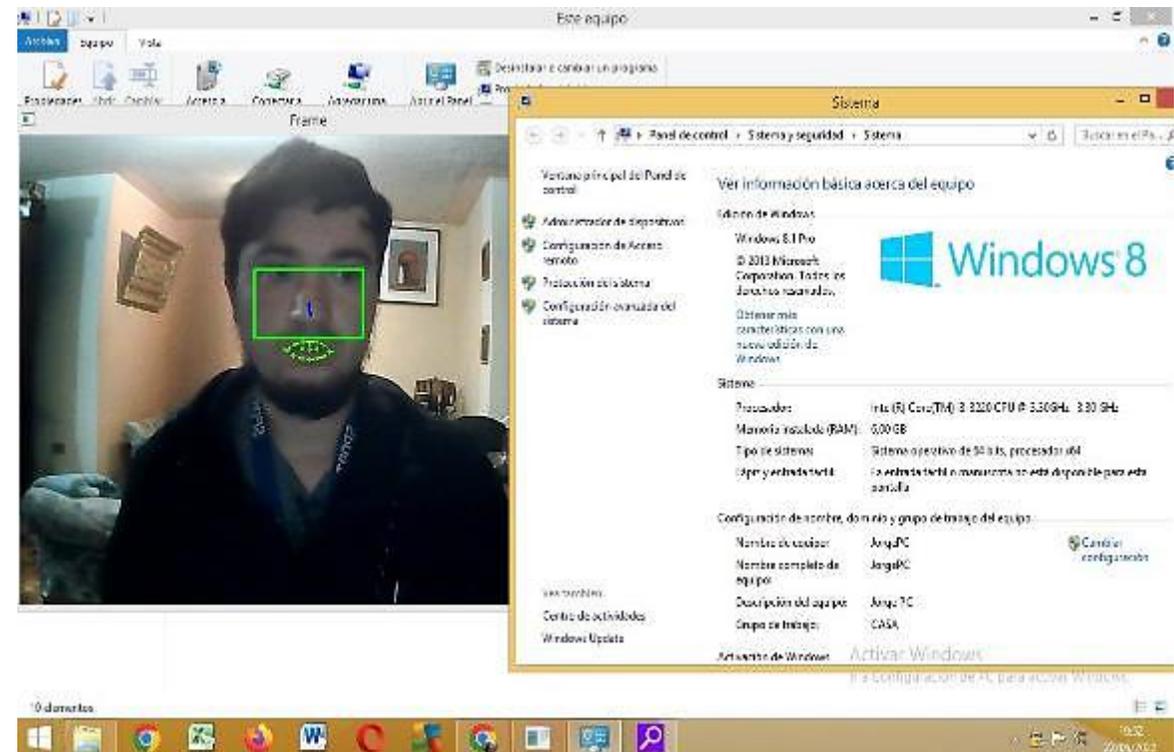


`C:\ProgramData\Microsoft\Windows\Start Menu\Programs\StartUp`

- Se probó el emulador en diferentes computadores de distintas gamas con diferentes características de hardware y software.
- Para este apartado se comparó solo computadoras con distintas versiones de Windows.
- Se buscó responder las siguientes preguntas para cada computador.
 - ¿Cuál es el costo computacional con el emulador activo?
 - ¿La velocidad de movimiento del cursor es la esperada?
 - Usando el micrófono más adecuado, ¿es rápida la respuesta del emulador a los comandos de voz?
 - ¿Funcionan correctamente todos los comandos de voz?
 - ¿Es necesario el uso de cámara externa?
 - ¿Es necesario el uso de micrófono externo?
 - ¿El emulador funcionó con normalidad?



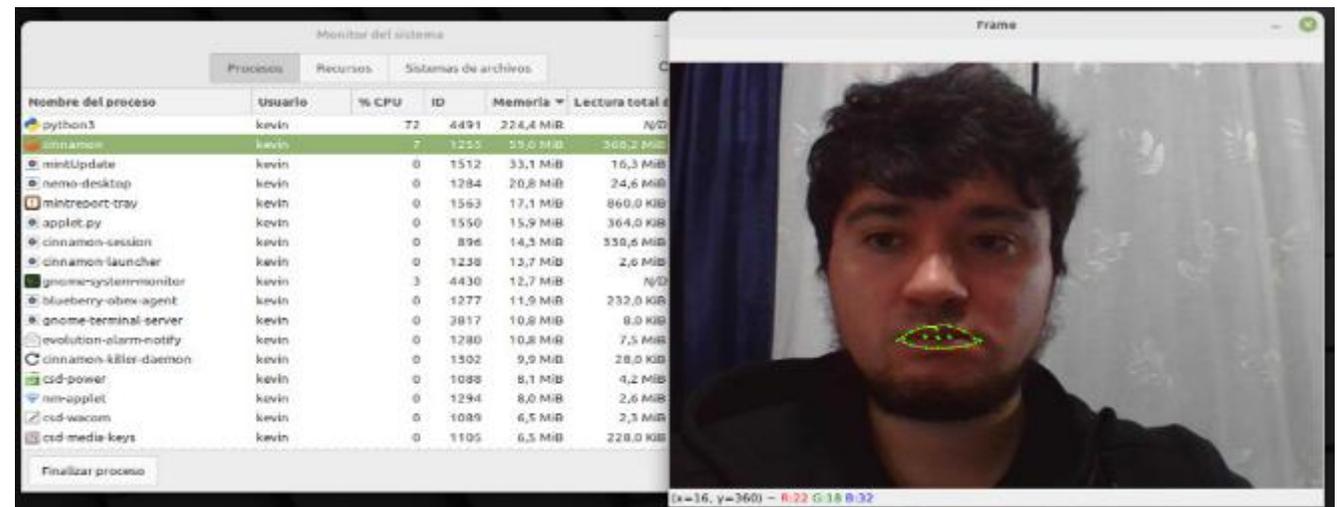
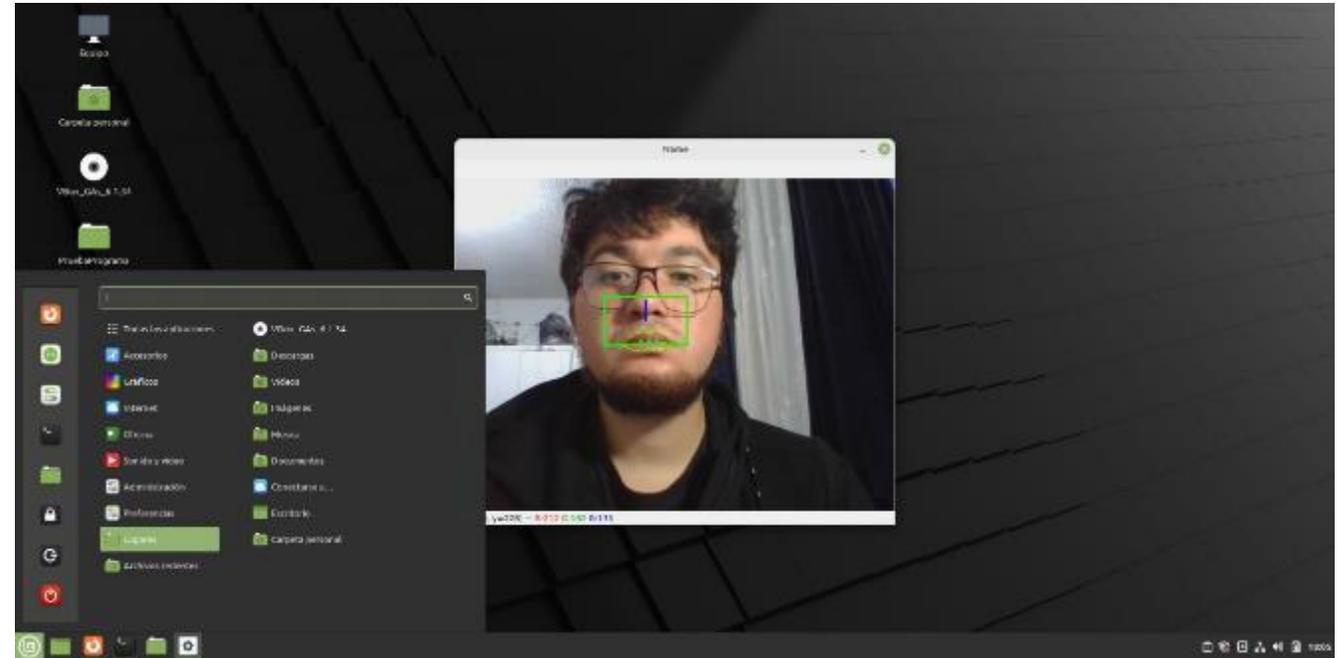
Nombre	Estado	CPU	Memoria	Disco	Red
Aplicaciones (4)					
Administrador de tareas		0,3%	11,9 MB	0 MB/s	0 MBps
Emulador_Mouvc		23,0%	280,3 MB	0 MB/s	0 MBps
Explorador de Windows		0,1%	98,1 MB	0,1 MB/s	0 MBps
Google Chrome		0%	55,5 MB	0 MB/s	0 MBps
Procesos en segundo plano (44)					
Adobe Acrobat Update Service L...		0%	11,1 MB	0 MB/s	0 MBps
Asistente de gráficos de disp...		0,4%	5,5 MB	0 MB/s	0 MBps
Aplicación de subsistema de col...		0%	10,3 MB	0 MB/s	0 MBps
AVG Antivirus Installer (32 bit)		0%	0,1 MB	0 MB/s	0 MBps
AVG Emergency Update (32 bit)		0%	0,2 MB	0 MB/s	0 MBps
AVG Overseer		0%	0,1 MB	0 MB/s	0 MBps
AVG Service (32 bit)		0%	22,0 MB	0 MB/s	0 MBps
AvLaunchComponent (32 bit)		0%	0,1 MB	0 MB/s	0 MBps



Máquina	Gama	Procesador	RAM	Sistema operativo	Costo Computacional	Velocidad del cursor esperada	Respuesta a comandos de voz	Funcionan Todos los comandos	Cámara externa	Micrófono externo	Novedad
Biostar H61MHB	baja	Intel Core i3-3220	6GB	Windows 8	23%	si	muy buena	Si	si	si	ninguna
Lenovo 81D6	baja	AMD A9-9425	8GB	Windows 10	47.8%	no	regular	Si	no	si	Computador saturado
HP – HP Laptop 15-db1xxx	baja	AMD Ryzen 3 3200U	8GB	Windows 10	24%	si	buena	Si	no	no	ninguna
Toshiba Satellite L 15W-B	baja	Intel Pentium N3540	4GB	Windows 10	-	-	-	-	-	-	El emulador no se pudo ejecutar.
Dell Inspiron 15 3515	media	AMD Ryzen 5 3450U	8GB	Windows 11	24.2%	si	muy buena	No	no	si	ninguna
ASUS Vivobook 15	media	Intel Core i5-8250	8GB	Windows 10	22%	si	muy buena	si	no	si	ninguna
Dell G3 3579	alta	Intel Core i7-8750H	16GB	Windows 10	15.4%	si	muy buena	si	no	no	ninguna
Toshiba Satellite P55W-C	alta	Intel Core i7-6500U	12GB	Windows 10	32.3%	si	muy buena	si	no	no	ninguna
Toshiba Satellite P75-A7200	alta	Intel Core i7-4700	8GB	Windows 8	13.8%	si	muy buena	si	no	no	ninguna
ASUS K53E	alta	Intel Core i7-2670	8GB	Windows 7	-	-	-	-	-	-	El emulador no se pudo ejecutar.

Pruebas en Linux Mint 20.3

- Se instaló el sistema operativo en una máquina virtual con 6GB de memoria RAM.
- El costo computacional de la aplicación en la máquina virtual fue alto (72%).
- En términos generales el emulador funcionó correctamente.
- La velocidad del movimiento del cursor fue un poco lenta debido al uso de la máquina virtual.



Pruebas en Raspberry Pi 3 B

- El costo computacional de la aplicación fue alto (52%).
- Esta computadora posee 1 GB de memoria RAM.
- El sistema se ejecutó correctamente.
- La velocidad de ejecución del emulador fue muy lenta debido a que las características del dispositivo son muy básicas



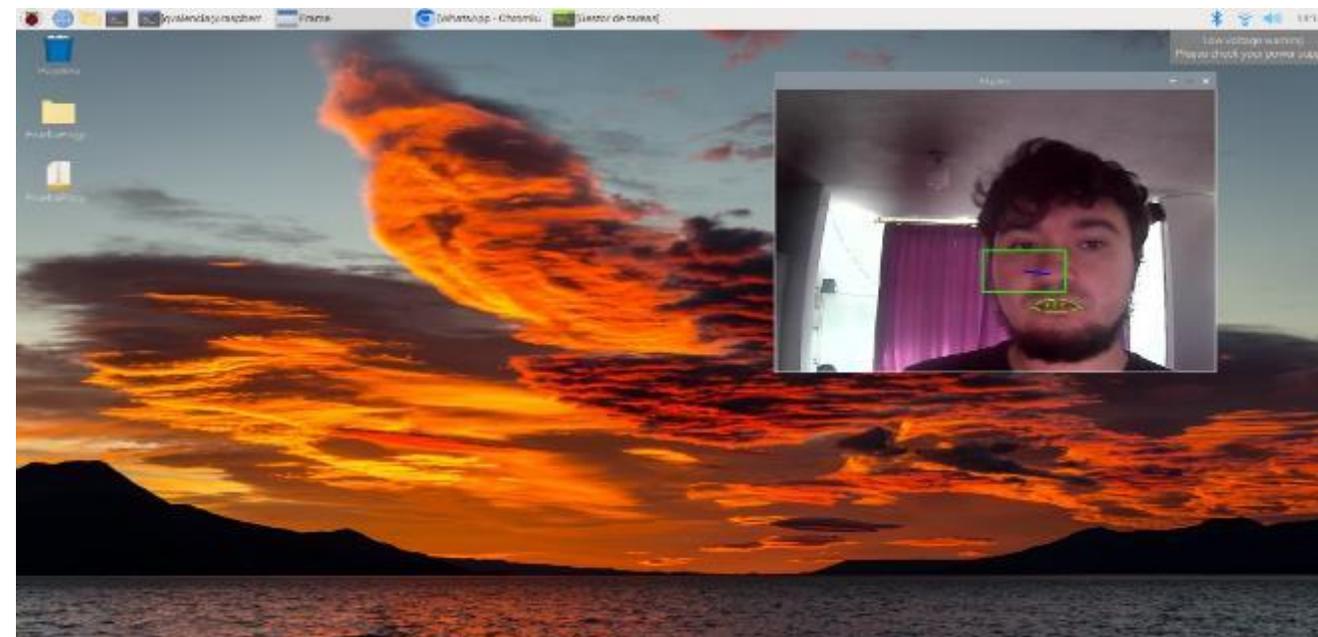
Gestor de tareas

Archivo Ver Ayuda

Uso de la CPU: 57% Memoria: 441 MB de 923 MB usados

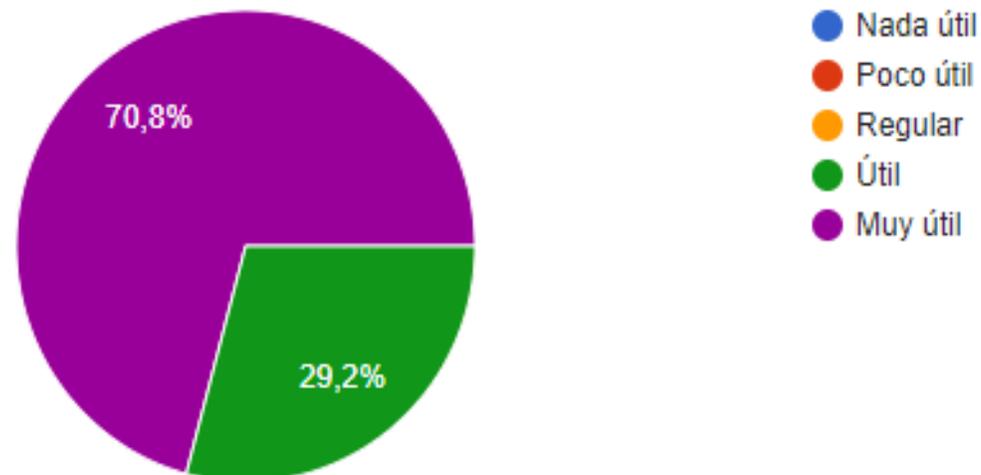
Orden	Usuario	% CPU	RSS	Memoria
python3	gvalencia	52%	252,3 MB	
pulseaudio	gvalencia	2%	7,6 MB	
lxtask	gvalencia	0%	19,5 MB	
chromium-browser-v7	gvalencia	0%	143,0 MB	
chromium-browser-v7	gvalencia	0%	80,1 MB	
chromium-browser-v7	gvalencia	0%	116,6 MB	
lxpanel	gvalencia	0%	21,0 MB	

más detalles Salir



- Con la finalidad de verificar la funcionalidad y perspectiva de los usuarios se puso a prueba el emulador con un grupo de 24 adultos de edades entre 22 y 82 años.
- El ejercicio de prueba consistió en realizar una consulta específica en el buscador de su preferencia.

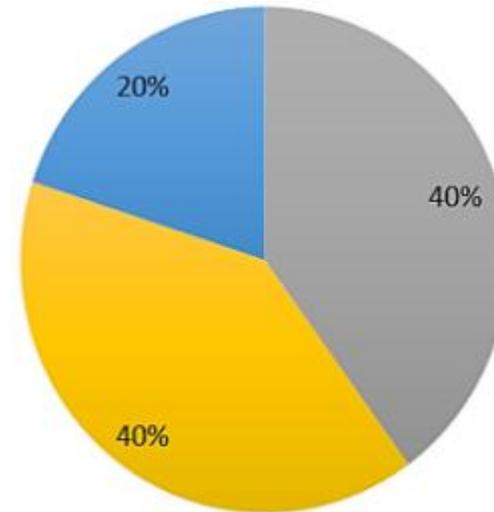
¿Qué tan funcional considera que es el emulador para personas con discapacidades motrices?



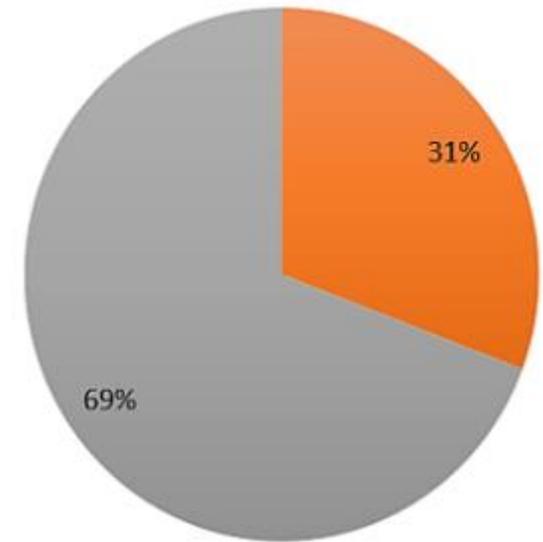
¿Qué tan difícil fue la adaptación al uso del emulador?



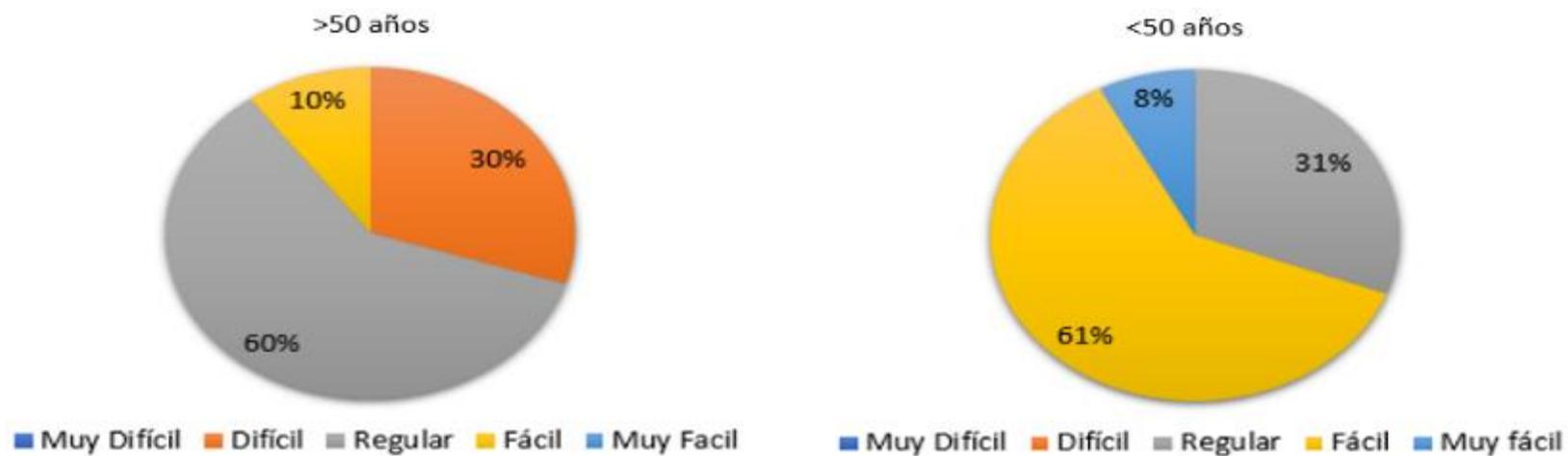
>50 años



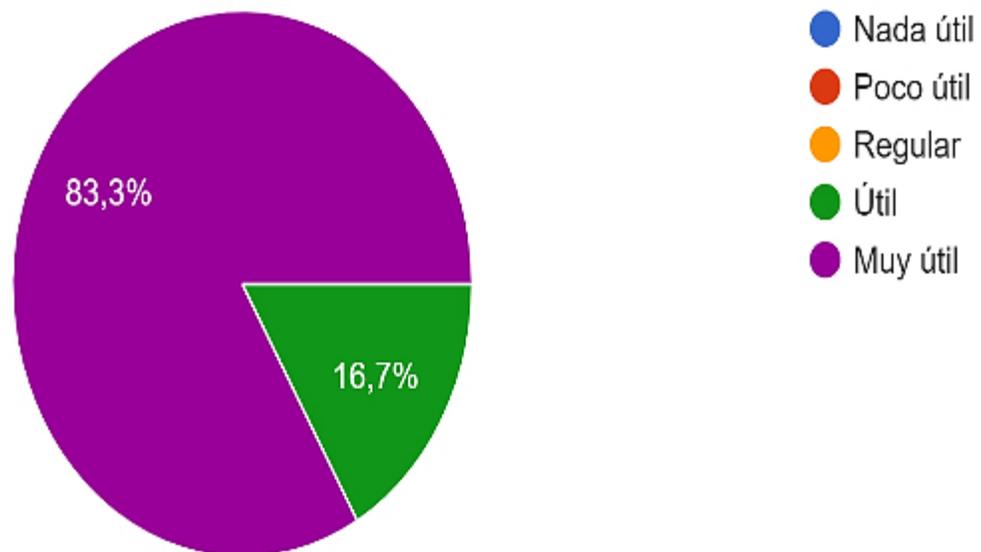
<50 años



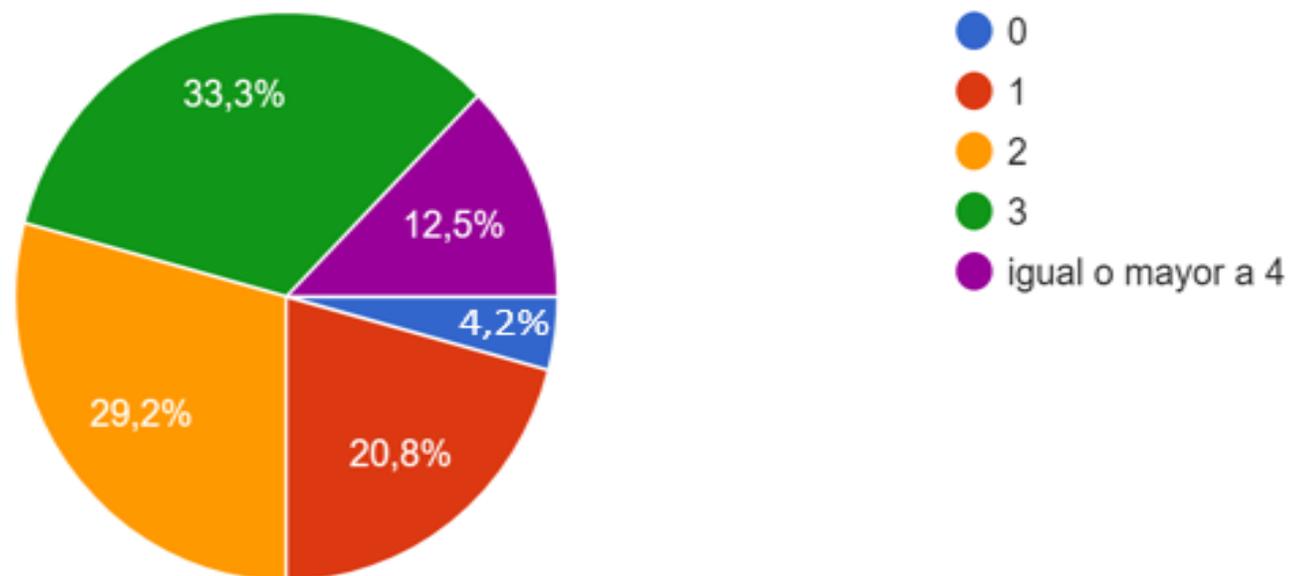
¿Qué tan difícil fue la realización del ejercicio?



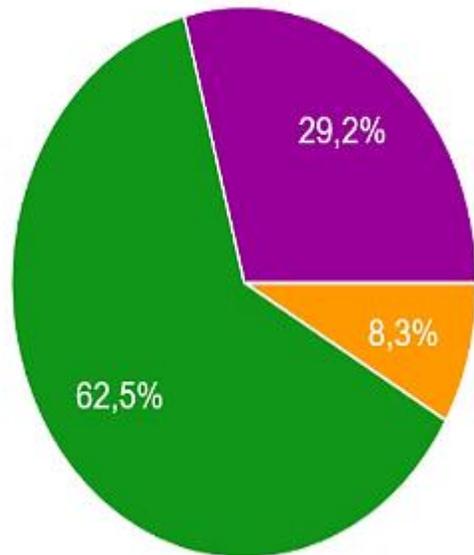
¿Qué tan útiles le resultaron los comandos de segmentación de pantalla?



¿Cuántas veces tuvo que reiniciar el movimiento del cursor (manteniendo abriendo la boca) para poder realizar el ejercicio?

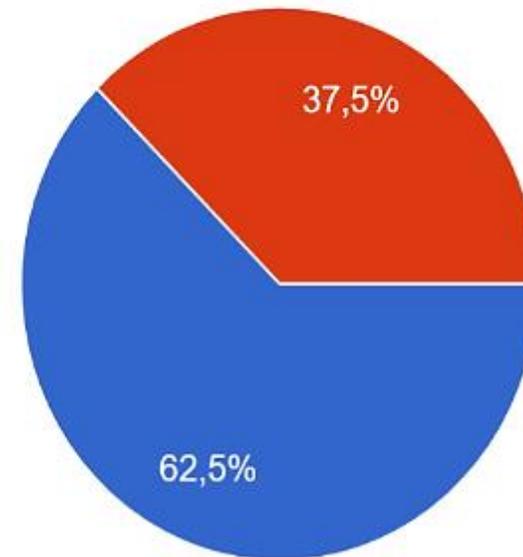


¿Qué tan preciso considera que es la detección de los comandos de voz en su computador?



- Muy mala
- Mala
- Regular
- Buena
- Muy buena

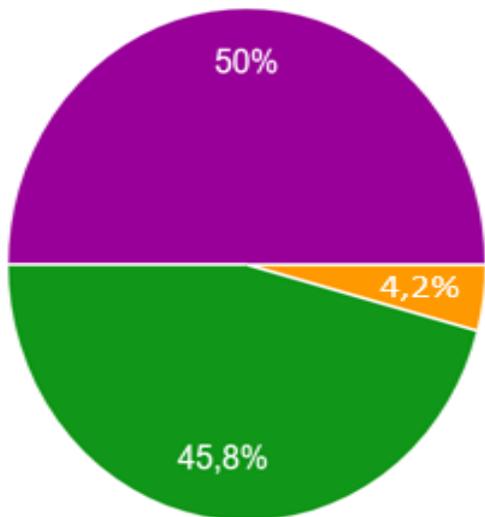
¿Utilizó un micrófono externo para realizar la prueba?



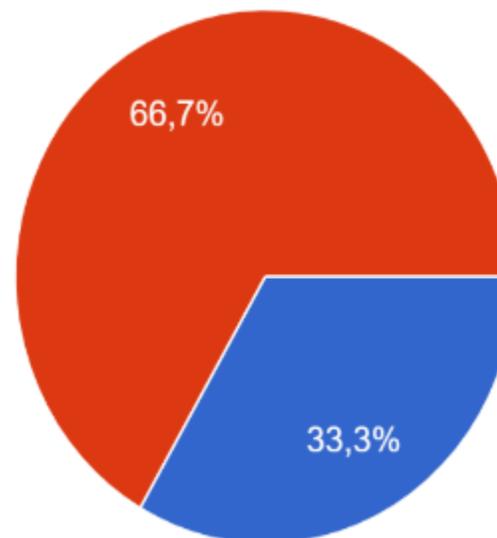
- Sí
- No

¿Qué tan preciso considera que fue la detección de su rostro?

¿Utilizó una cámara externa para realizar la prueba?



- Muy mala
- Mala
- Regular
- Buena
- Muy buena



- Sí
- No

Cronometrar el tiempo que tardó en realizar el ejercicio

- A partir de las respuestas recibidas, se obtuvo un tiempo promedio de 44.04 segundos.
- Se realizó pruebas del mismo ejercicio, pero usando un mouse físico y tipeando a velocidad moderada las palabras con el teclado.
- Se cronometró un tiempo promedio aproximado de 12.4 segundos

Si bien es una gran diferencia, se debe considerar que el emulador está enfocado para personas con discapacidades motrices, por lo que no se espera que el desempeño a la hora de manejar el cursor con el emulador, sea mejor que cuando se usa un ratón común.



Conclusiones



- En este trabajo se realizó el diseño, la implementación y la validación de un sistema que permite emular las funcionalidades de un mouse, a través del uso de algoritmos de visión artificial y reconocimiento de voz.
- El uso del emulador no sería viable en aplicaciones que requieran un uso muy fluido del cursor, como puede ser el caso de videojuegos complejos.
- En cuanto al modo dictado que se implementó, su uso no sería lo suficientemente fluido para poder redactar con facilidad documentos de texto formales
- Se realizó pruebas de funcionamiento del emulador variando la iluminación y se determinó que el sistema puede funcionar en entornos poco iluminados, sin embargo, dependerá en gran parte de que tan bien se logre enfocar e iluminar el rostro del usuario.

- Se realizaron pruebas de desempeño en varios computadores de diferentes características y se determinó que el emulador puede funcionar en dispositivos de todas las gamas. Esto permite concluir que el emulador cumple con el requerimiento de ser de bajo costo ya que puede funcionar en máquinas de gama baja, el único gasto adicional que representaría sería la adquisición de una cámara o micrófono externo, si fuese necesario.
- El emulador puede funcionar de buena manera en máquinas con sistemas operativos de distribución libre como Linux.
- Basándose en las pruebas de desempeño en varios computadores, se pudo determinar que las características mínimas del computador donde se ejecute el programa deberán ser: un procesador igual o mejor a un i3, una memoria RAM mayor a 6GB y sistema operativo superior a Windows 7.
- Se logró validar la funcionalidad del emulador a través de pruebas a distintos usuarios, quienes a través de una encuesta consideraron que puede ser de gran utilidad para personas que padezcan de discapacidades motrices.

- Se realizó un análisis para determinar cómo es el funcionamiento de la detección de comandos de voz y de los rostros los cuales dieron resultados positivos. También se les preguntó si se realizó la prueba con dispositivos externos, y los resultados mostraron que la mayoría de usuarios no utilizó una cámara externa, sin embargo, si utilizaron micrófonos externos. Esto indica que los requerimientos de hardware para lograr un buen desempeño del reconocimiento de rostros son menores a los necesarios para una buena detección de la voz.
- En las pruebas a usuarios se les solicitó su opinión acerca de características específicas del emulador, como: la velocidad de movimiento, los comandos de segmentación de pantalla. El 100% de los usuarios se sintió a gusto con la velocidad a la que se mueve el cursor y esto se complementó con los resultados positivos de la pregunta referente a la utilidad de los comandos de segmentación de pantalla, los cuales se desarrollaron con la finalidad de permitir un movimiento más rápido.



Referencias

- Mosquera, J., Loaiza, H., Nope, S., & Restrepo, A. (2020). Human-computer multimodal interface to internet navigation. *Disability and Rehabilitation: Assistive Technology*, 16(8), 807–820. <https://doi.org/10.1080/17483107.2020.1799440>
- Rosebrock, A. (2017, abril 3). *Facial landmarks with dlib, OpenCV, and Python*. PyImageSearch. <https://pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/>
- Ferrin, C., Mosquera, J., Pino, J., Moctezuma, L., Burgos, J., Aragón, L., & Loaiza, H. (2021). Interfaz humano-computador basada en gestos faciales y orientada a la aplicación WhatsApp para personas con limitación motriz de miembros superiores. *TecnoLógicas*, 24(50), 72–96. <https://doi.org/10.22430/22565337.1722>
- Díaz, J., & Morales, C. (2019). *Desarrollo de una interfaz mediante señales EOG para el manejo de la computadora por parte de una persona con discapacidad en los miembros superiores*. [Universidad Peruana de Ciencias Aplicadas]. https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/652754/D%C3%ADaz_SJ.pdf?sequence=11&isAllowed=y
- Chandra, A. (2018, octubre 7). *Mouse Cursor Control Using Facial Movements — An HCI Application*. Towards Data Science. <https://towardsdatascience.com/mouse-control-facial-movements-hci-app-c16b0494a971>
- Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2013). 300 faces in-the-wild challenge: The first facial landmark Localization Challenge. *Proceedings of the IEEE International Conference on Computer Vision*, 397–403. <https://doi.org/10.1109/ICCVW.2013.59>