InteractiVue

Control the video player

with your face

WS 23/24 Interaction Engineering

Prof. Dr. Michael Kipp

Augsburg University of Applied Sciences

Abdelmonem Abbas (2171832)

Daria Kazmina (2193621)

Table of Contents

Abstract

The InteractiVue project introduces an innovative hands-free video player that responds to facial gestures, revolutionizing user interaction with video content. By interpreting facial expressions, including mouth puckering, eyebrow movements, open mouth, and compressed lips, the system empowers users to control actions like play, pause, volume adjustment, and video navigation without relying on traditional mouse or hand-based gestures. This breakthrough technology not only enhances the accessibility and convenience of video playback but also holds potential benefits for individuals with disabilities and those navigating scenarios where hands-free control is paramount.

Motivation

The primary objective of this project is to investigate the feasibility of controlling a video player using facial gestures. Our research aims to identify the most user-friendly interactions for managing various video player functions and explore the potential correlation between distinct facial gestures and their applicability to specific video player features.

A crucial aspect of our motivation is to extend the benefits of this technology to individuals with disabilities, offering an alternative means of video player control for those who may face challenges with conventional methods like mouse and keyboard interactions. Additionally, we consider the practical context of individuals with occupied hands, such as someone engaged in activities like cooking, where hands-free control becomes especially valuable.

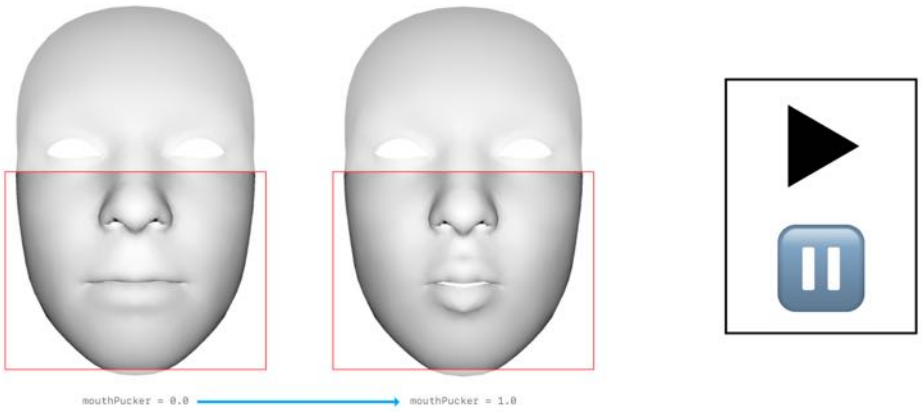
By delving into the realm of facial gesture recognition for video player interaction, this project seeks to enhance accessibility and convenience, providing a valuable tool for a diverse range of users, including those with disabilities and those navigating situations where hands-free control is essential.

Concept

In the process of conceptualizing the gesture-based interaction design for the hands-free video player, our primary objective was to identify intuitive gestures seamlessly integrated with the video player's functionalities. During the prototype development phase, we encountered a significant challenge: MediaPipe, the face detection library we utilized, simultaneously analyzed multiple facial gestures. Through thorough experimentation, we strategically curated a selection of gestures that MediaPipe could accurately detect, driven by a logic deeply rooted in the relationship between facial expressions and corresponding video player actions.

Play / Pause: Mouth Pucker

The deliberate act of mouth puckering was chosen to activate both the play and pause functions. This selection is grounded in the logic that the motion of puckering one's lips mirrors the action of clicking a button—an instinctive and natural movement akin to initiating or pausing video playback.

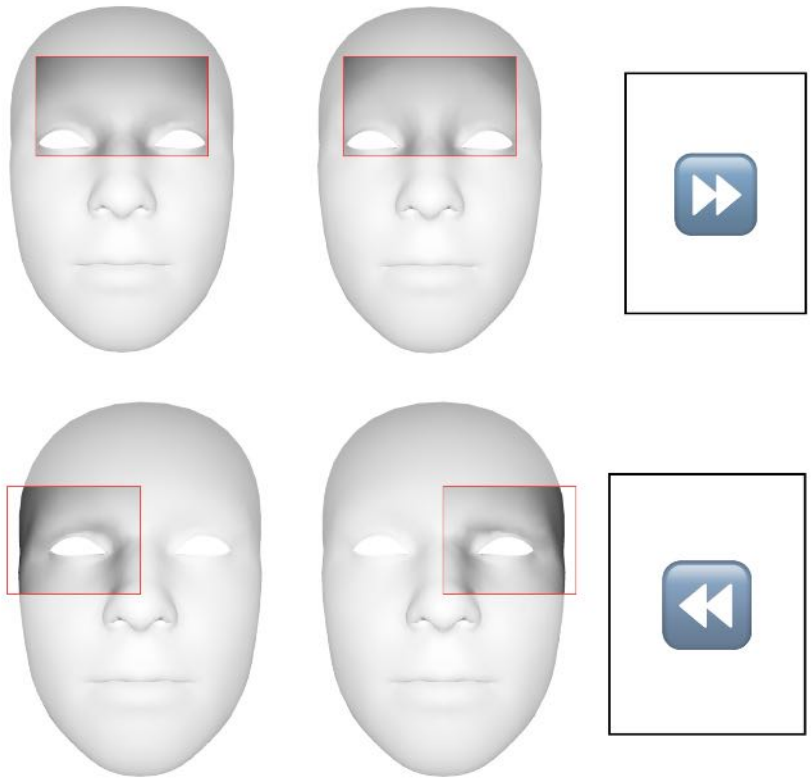


Moving Forward / Moving Backward: Eyebrows Up & Eyebrows Down

Implementing eyebrow gestures for video navigation required a nuanced and responsive approach. Raising eyebrows triggers a forward progression of the video by 10 seconds. Crucially, this forward motion persists as long as the user maintains raised eyebrows, offering a seamless and continuous exploration, akin to looking ahead while navigating content.

Conversely, lowering eyebrows initiates a backward movement, retracing content for 10 seconds. Similar to forward motion, the backward progression continues as long as the user keeps their

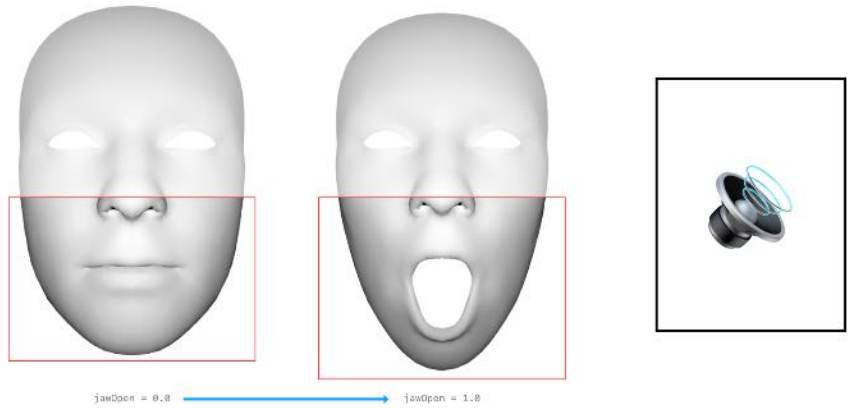
eyebrows lowered until they reset their eyebrows. This design choice aligns with the intuitive connection between the gesture and the desired action, providing users with a dynamic and responsive means of navigating through video content.

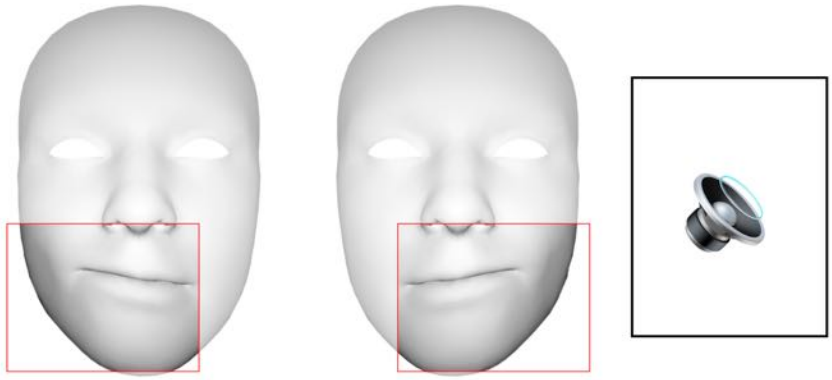


Volume Up / Volume Down: Open Mouth & Compressed Lips

The volume control functions are intricately tied to mouth movements, leveraging the inherent relationship between facial expressions and sound. Opening the mouth prompts a gradual increase in volume, raising it by 10%. Importantly, this volume increase persists as long as the user sustains an open-mouth gesture, offering a continuous and customizable audio experience. This design aligns with the natural inclination to amplify sound by opening one's mouth, providing an intuitive and responsive means of adjusting audio levels.

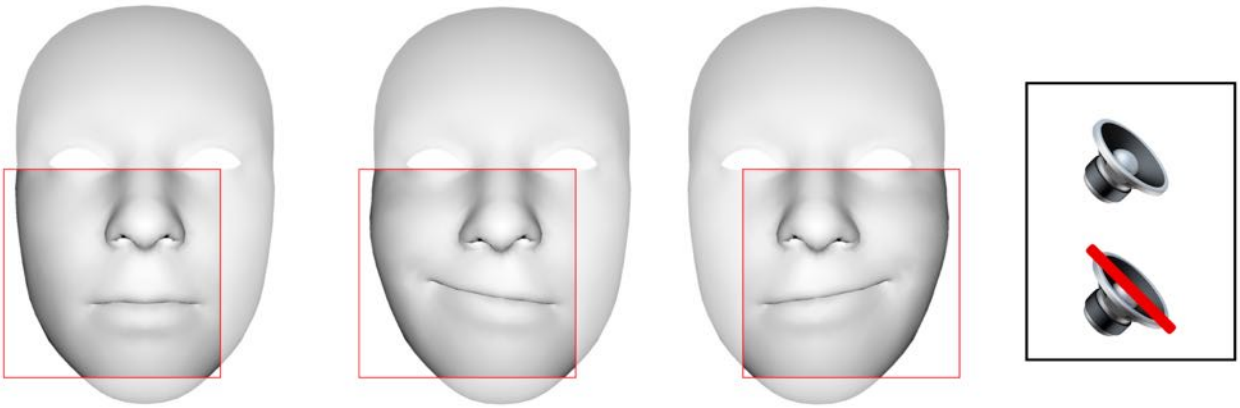
Conversely, compressing the lips signals a desire for decreased volume. The volume decreases by 10%, reflecting the instinct to reduce sound by closing one's mouth. Similar to volume increase, the volume decrease continues as long as the user maintains the compressed lip gesture until they choose to reset their lips. This thoughtful integration of gestures not only establishes an intuitive connection between facial expressions and volume control but also ensures a dynamic and user-friendly audio adjustment experience within the hands-free video player.





Mute / Unmute: Smile

The act of smiling serves as a distinctive gesture to mute or unmute the video. This choice is rooted in the logic that a smile, a universally recognized expression of happiness or approval, is associated with the concept of sound being either enabled or disabled—a logical extension of the relationship between facial expressions and audio control.



The interactions involving mouth press and smile strategically track the left and right sides of the lips. This deliberate design choice simplifies the execution of the full action, prompting users to

perform the entire gesture on just one side of their face to reach the trash-hold value. While this may initially pose a challenge, it aligns with the natural asymmetry of facial movements, intentionally making it more difficult to execute completely symmetrical actions with both sides of the face. This consideration adds a layer of realism to the gesture-based interaction design, acknowledging the inherent asymmetry of facial expressions.

Furthermore, the extensive list of motions in the MediaPipe library presented a challenge due to potential overlap between gestures. To address this, we meticulously identified gestures that minimize interference, ensuring that triggering one gesture does not unintentionally activate another. Careful consideration was given to gestures like eyebrow movement, cheek motion, and nose gestures, aiming to isolate distinct actions without triggering unintended responses. This strategic approach enhances the precision and reliability of the hands-free video player, offering users a seamless and intuitive interaction experience. The thoughtful alignment of gestures with minimized interference ensures a smooth and natural interaction flow, enriching the user experience.

Implementation

This section provides a detailed account of the developmental journey undertaken in creating the interactive hands-free video player project, shedding light on challenges faced and innovative solutions implemented.

The initial phase of the project involved leveraging Python's extensive library ecosystem, with a focus on Dlib and OpenCV for facial landmark detection. A custom video player, utilizing VLC libraries, was also crafted. Despite the promise of Python's versatility, two major roadblocks emerged. Firstly, the face recognition system encountered notable delays and lag, significantly impacting user experience. Secondly, the integration of sound playback through VLC presented compatibility issues, necessitating an update that was yet to be released. Faced with these challenges, a strategic decision was made to transition towards a web-based paradigm.

Motivated by the need for improved performance and cross-platform accessibility, the project shifted towards a web-based application. The chosen technology stack aimed to harness the strengths of various tools:

- **MediaPipe:** Google's framework provided pre-trained models for facial landmark detection, substantially expediting development efforts.
- **TypeScript:** Renowned for its strong typing and object-oriented features, TypeScript ensured code maintainability and scalability, aligning with the project's evolving requirements. It's important to note that TypeScript, being a superset of JavaScript, requires compilation to JavaScript before running on the web.
- **HTML/CSS:** These foundational web development languages played a crucial role in constructing an aesthetically pleasing user interface for the video player.

Integration Highlights:

- **Face Recognition:** The integration of MediaPipe's Face Landmarker model played a pivotal role in tracking facial landmarks, facilitating the interpretation of user gestures such as mouth puckering and eyebrow movements.
- **Video Player:** A bespoke video player, built from the ground up using web technologies, ensured seamless playback and resolved sound compatibility issues.

- **User Interface:** Meticulous use of HTML and CSS was employed to craft an intuitive and visually appealing interface for controlling video playback.

In the developed system, the triggering of functions operates on a seated trash-hold mechanism. For instance, when a facial gesture, like mouth puckering, reaches a predefined value (e.g., 0.9), it serves as a trigger to activate a specific function, such as clicking the play/pause button. The time lag, however, is designed to be short enough to allow for continuous triggers. For example, holding a gesture, such as an open mouth for volume increase, enables a continuous activation of the associated function. This innovative approach empowers users to control various aspects of the video player seamlessly by holding specific gestures for as long as needed, adding a layer of intuitiveness and adaptability to the hands-free video player experience.

Evaluation

Our evaluation aims to pinpoint design issues, enhance the project, and understand user preferences regarding the developed interaction methods. To streamline the evaluation process, we focused on testing three specific gestures, mouth pucker, smile, and eyebrows, for video play and pause functionality. This targeted approach allowed us to assess the effectiveness of these gestures and gather valuable insights for potential improvements.

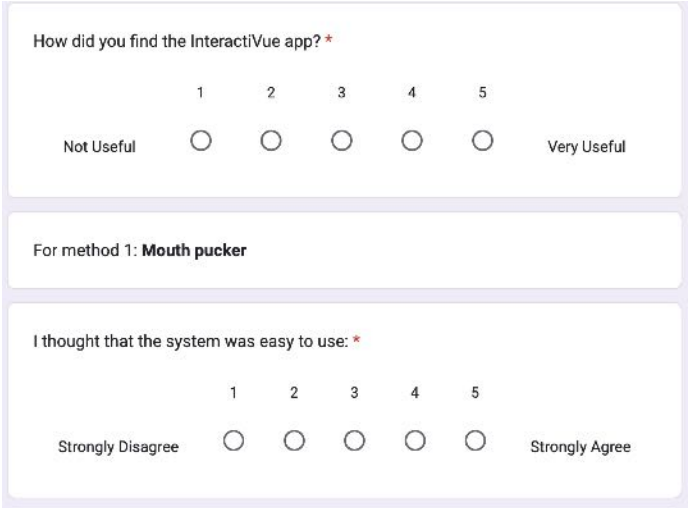
User test

We conducted an online user test involving six participants (one woman aged 23-30 and five men aged 25-32). The test followed this sequence:

1. Users received an overview of the video player concept and instructions on engagement.
2. The video clip, featuring a butterfly, prompted users to pause and play whenever the butterfly appeared. This task was executed using each of the three gestures.
3. After engaging with the gestures, users completed a questionnaire through an online form.

The Questionnaire

The post-test questionnaire aimed to gauge users' confidence levels, opinions, and preferences for each method. Key questions included:



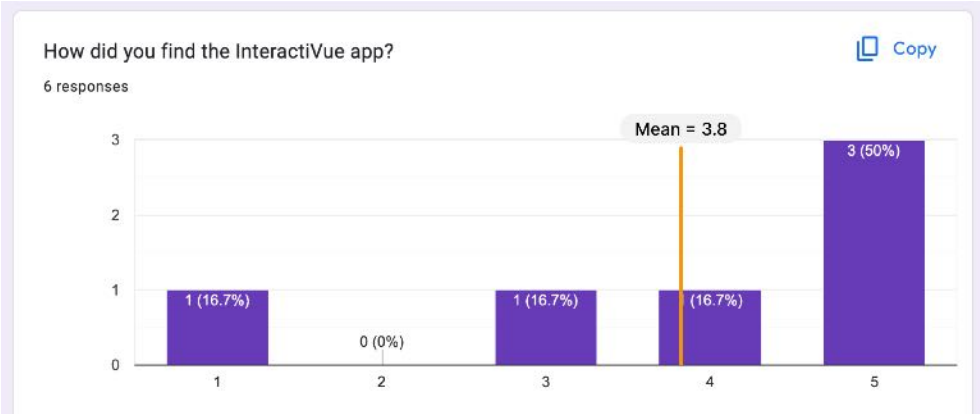
- "**How did you find the InteractiVue app?**" We employed a 5-point Likert scale, ranging from "Not Useful" to "Very Useful," to assess the overall value of the solution for users.
- For Each Method: "**I thought that the system was easy to use.**" "I **felt very confident using the system.**" "**I think that I would use this system frequently.**" These questions, rated on a 5-point Likert scale, aimed to measure ease of use, user confidence, and future usability expectations.

At the end of the survey:

- "**What is your favorite method to play/pause the video**?" Users were provided with the three choices: Mouth Pucker, Smiling, and Eyebrows, to identify their preferred interaction method.

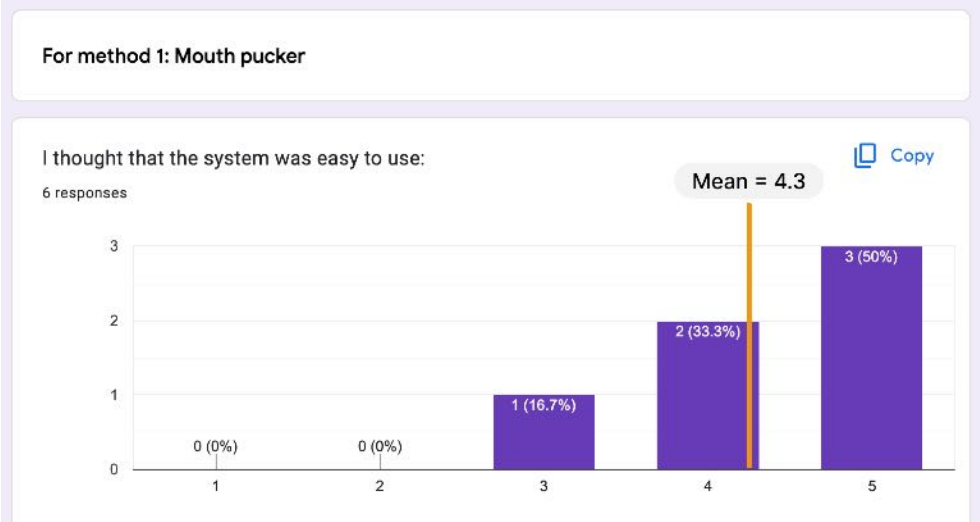
Results

Rating of InteractiVue App: The majority of users (50%) found the InteractiVue app very useful, with a mean rating of 3.8.

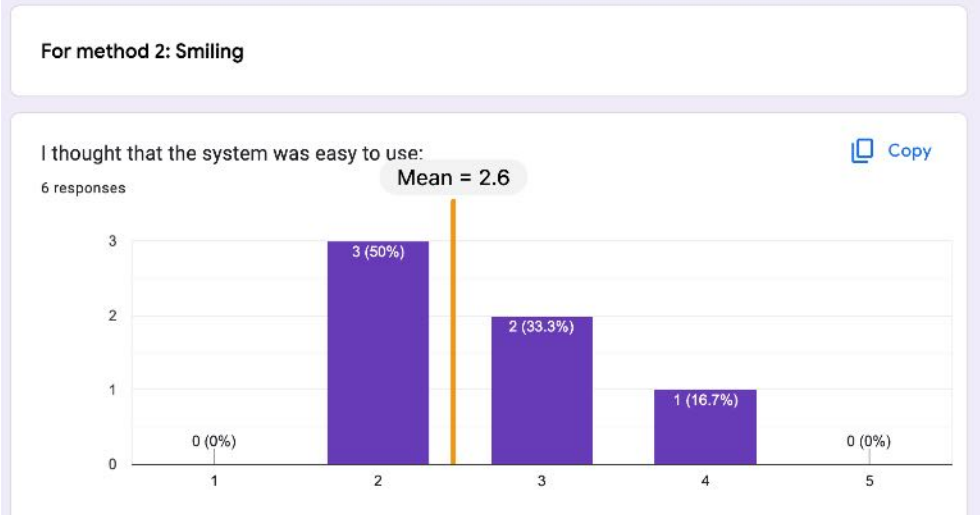


For each method we asked about how was it easy to use, and looking for results we will find that the:

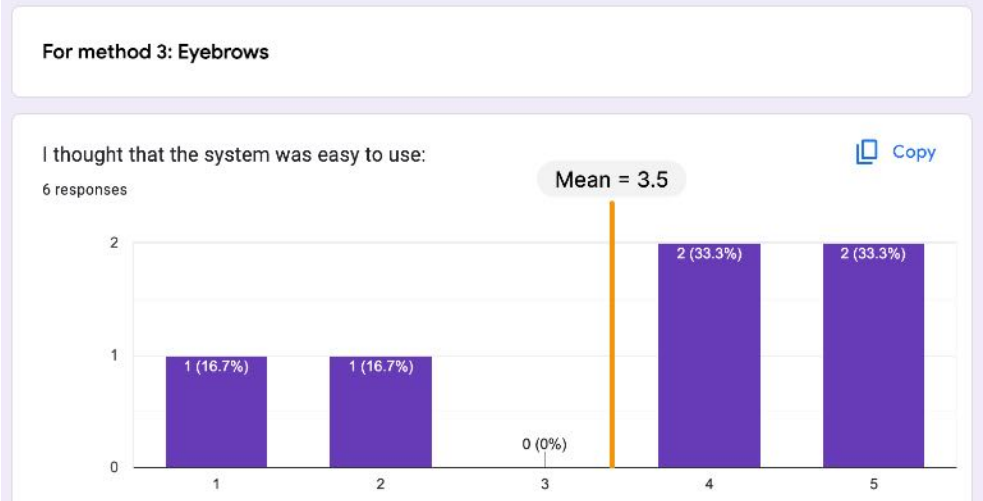
Mouth Pucker (mean = 4.3): Users perceived the Mouth Pucker as the easiest gesture, requiring minimal effort due to its uniqueness.



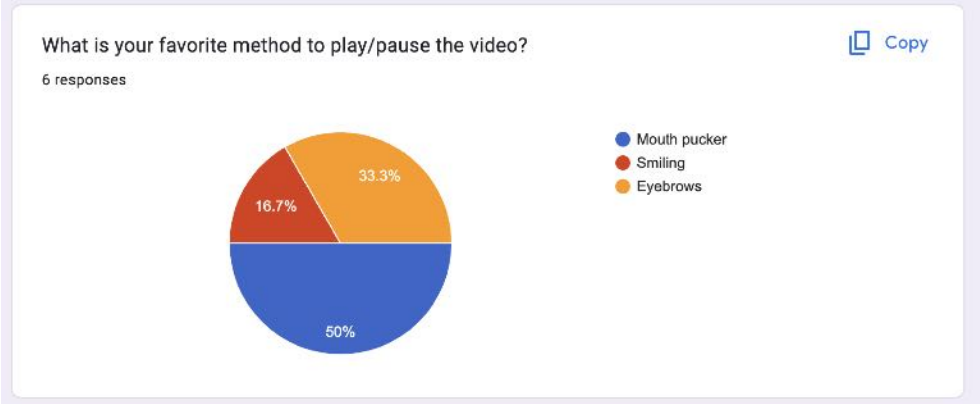
Smiling (mean = 2.6): While the Smiling gesture was considered easy, some users found it slightly annoying, unintentionally smiling during video playback.



Eyebrows up (mean = 3.5): The Eyebrows Up gesture was viewed as unique but tiring, requiring conscious effort to maintain.



At the end of the survey, we asked the main question: “What is your favorite method to play/pause the video?”, and 50% of users favored the Mouth Pucker gesture, citing its ease and sustainability.



Findings

Post-test feedback provided valuable insights:

- Users preferred the Mouth Pucker gesture.
- Initial attempts involved slower puckering, with users adapting to a faster motion.
- Smile gesture was less effective for those who involuntarily kept smiling.
- Eyebrows gesture worked well for users wearing glasses, emphasizing its suitability for diverse user scenarios.

Conclusion

InteractiVue has proven to be a valuable addition, receiving favorable reviews for its user-friendly approach. Employing facial gestures for video player control has garnered positive feedback, particularly benefiting individuals with occupied hands or disabilities that impede traditional mouse and keyboard interaction. Looking ahead, our focus is on refining user interactions for improved simplicity and reduced intrusion. Future explorations may involve alternative facial recognition libraries to enhance precision. Despite its current success, we are committed to ongoing improvement, aiming for a more streamlined and user-centric hands-free video player.

Future work

As we envision the future trajectory of our project, several avenues for improvement and expansion emerge. The evolution of our hands-free video player could encompass the following enhancements, shaped by user feedback and ongoing innovations:

Enhanced Front-End Design: the aesthetic appeal and user interface (UI) design of our video player represent crucial facets that warrant refinement. Future iterations could focus on enhancing the front-end to provide a visually pleasing and immersive user experience. Incorporating modern design principles and ensuring a user-friendly interface will contribute to the overall attractiveness and usability of the hands-free video player.

Alternative Mute/Unmute Interaction: while the smile gesture serves as an intuitive means to mute or unmute the video, exploring alternative interactions for this specific function is an avenue for improvement as a smile is a natural human reaction which can lead to false triggers. Understanding that different users may have varied preferences, we aim to investigate alternative gestures or controls that offer flexibility and cater to a diverse range of user needs. This exploration seeks to enhance the adaptability and inclusivity of our hands-free video player.

Development of a Full-Stack Application: to elevate the user experience and extend functionality, future efforts may involve transforming the hands-free video player into a comprehensive full-stack application. This evolution could empower users to have a more holistic interaction with the system, enabling them to initiate their video playback experience by selecting content from an interface. Integrating a full-stack architecture would involve back-end development to handle video selection, user profiles, and other backend functionalities, complementing the existing front-end features.

Advanced Gesture Recognition: exploring and implementing more sophisticated gesture recognition capabilities is another promising avenue for future work. This could involve fine-tuning existing gestures, introducing new ones, or leveraging advanced machine learning models to enhance the precision and diversity of recognized facial interactions. Advancements in gesture recognition technology would contribute to a more seamless and responsive user experience.

Integration with External Devices: consideration may be given to integrating our hands-free video player with external devices, expanding its compatibility and usability. This could involve

synchronization with voice-activated assistants, smart home devices, or other emerging technologies, providing users with an even more interconnected and versatile media consumption experience.

Accessibility Features: to further support users with diverse needs, incorporating additional accessibility features could be explored. This might involve options for customizable gestures, voice commands, or other assistive technologies to ensure that the hands-free video player remains an inclusive tool for a broad user base.

In conclusion, the future development of our hands-free video player is envisioned as a dynamic process of continual improvement and innovation. By embracing user feedback, refining design elements, exploring alternative interactions, expanding application capabilities, and considering the integration of advanced technologies, we aim to create a hands-free video player that is not only technologically advanced but also highly adaptable and user-centric.