Face Media Player: Use your face to control a media player

Diwas Lamsal
st122324@ait.asia
Asian Institute of technology
Thailand

Harold Popluhar
st122556@ait.asia
Asian Institute of technology
Thailand

Ayush Koirala
st122802@ait.asia
Asian Institute of technology
Thailand

ABSTRACT

As the Computer Vision techniques have evolved, people are coming up with novel ways to use face-based interaction techniques which provide several benefits, particularly in accessibility. Many recent works have presented novel ways for text input or as unique interaction methods in video games. It might also benefit to include such techniques in media players as the video consumption is at about an all-time high. To our knowledge, these works have not (or only partially) covered navigation methods for a media player, which is among the most popular applications. In this paper, we compare some face-based hands-free interaction techniques to see how they perform for controlling a media player. For doing this, we developed a media player that allows interaction using nose tracking, wink, blink, mouth open, and eyebrow movement. From a controlled user study with 12 participants, we found that nose tracking for volume control and video navigation, and mouth open for pause/play are the best combinations for controlling our media player.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Hands-free Input; Gesture Interaction; Facial Gestures; Nose; Blink; Mouth; Eyebrows.

ACM Reference Format:

Diwas Lamsal, Harold Popluhar, and Ayush Koirala. 2022. Face Media Player: Use your face to control a media player. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Input modalities using voice, gesture (face, hand, etc.), and tracking (head, hand, gaze, etc.) have recently started getting more attention. Even 20 years ago, they have proven to have great potential as alternative interaction methods for people with severe disabilities [2, 15]. With today's improvements in cameras, computation power, and deep learning, these interaction methods have also improved considerably. A number of recent works have presented novel use of such interactions and show promising results [1, 9, 11, 14, 27, 28]. Such interactions might be the best viable alternatives for hands-free interaction [21] (in Hands-free interaction, a user should be able to interface with the computer without the use of their hands [29]). [21] shows that voice is the most studied and a good-performing hands-free interaction technique. Other relatively less-studied, but also viable hands-free interaction methods utilize facial features. Computer Vision algorithms have matured and brought significant improvements in camera-based interactions like hand and facial gestures [5, 20, 23]. Some have combined multiple input methods like touch and blink [28], and also compared them with traditional methods like keyboard and tilt (for smartphones) [1, 9]. Papers like [27] and [14] have implemented gaze-based inputs as novel game interaction techniques, and also highlight how modern games like Battlefield, Far Cry, and Assassin's Creed have included gaze-based interactions as part of gameplay. Other modalities such as Brain-Computer Interface are also used in some occasions [21], but they require a separate device such as an Electroencephalogram (EEG), which might not be accessible to most people.

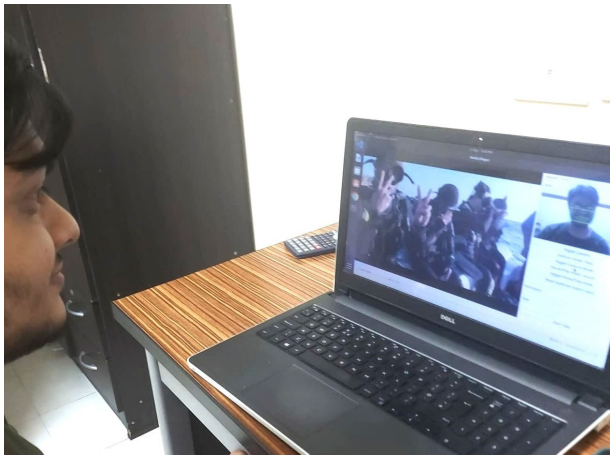

Figure 1: Experimental setup. Participants used face gestures and nose tracking to navigate the video; perform play, pause; and volume control.

Hands-free interaction techniques might come handy for people with certain diseases such as Arthritis, disabled people without the ability to use their hands, or also normal people (some scenarios include having their hands busy, hand fatigue, injury, posture, etc. [25]). We found voice to be the most commonly used technique for such interactions [21]. We also found that facial gestures are not as common despite performing very well with disabled people [20].

Conference acronym 'XX, June 03–05, 2022, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

Modalities such as face-gestures could be the only feasible hands-free interaction technique for a person who cannot speak. Most of the literature about facial gestures are implemented for text input and games. The common way to navigate a media player (which is one of the more widely used applications as video consumption has grown rapidly [26]) is through mouse, keyboard, or a touch screen. While few papers such as [7] have looked at using voice for interacting with videos, to our knowledge, we found no face-based methods for controlling a media player.

This paper presents hands-free interaction methods for a media player, followed by a controlled user experiment to assess these methods' performance. In particular, we attempt to evaluate the performance of using facial-gestures and nose tracking to control a media player. The key questions we are trying to address are: Is it possible to use face-based interaction techniques for controlling a media player with complete hands-free interaction? Can this method result in good performance and the overall perceived ease of use? What would be a suitable combination of such facial gestures for complete control of a media player? For answering these questions, we designed a prototype media player that allows face-gesture based inputs. We used mouth, eyebrow, wink, and nose tracking for performing the pause/play, volume up/down, and forward/backward operations in the media player. For controls, we follow the guidelines from Hands-free for Web [24] on playing media content and map face-based interactions to a subset of commands listed here.

In our prototype media player, for pause and play, we adapted two gestures: mouth open and blink; for navigation, which involves moving the video forward and backward, and increasing or decreasing the volume, we used a combination of two different techniques: tracking nose position (left, right up, and down), and wink (left and right) + eyebrow movement (up and down). We conducted an experiment with 12 participants, where we compared these different combinations with each other for two different lengths of videos. We found that the nose tracking + mouth open gesture performed the best among all the combinations. The participants found this combination to have less mental and physical workload, easier and enjoyable to use, and with a better perceived accuracy. We also found nose horizontal movement significantly outperforming wink for forward and backward commands in terms of task completion time.

Our contributions in this paper include:

- We provide an open-source media player prototype that allows interactions using different combinations of facial gestures.
- We compare different face-based interaction methods to evaluate which would be the most suitable combination for the tasks involved in controlling a media player in terms of the task workload, completion time, and preferred method. The results come from a controlled user study.
- We advance Human-Computer Interaction research by exploring novel ways to interact with a media player using only a person's face.
- Based on our results, future work can explore the use of such face-based interaction methods in other niche use cases such as a PDF reader

2 RELATED WORK

Most of the work in face-based interactions have dealt with text input, navigation and selection, and video games. In many cases, facial gestures seem to be used in combination with gaze or head tracking. In [12], they use gaze tracking and head tracking (tracking head movement using the nose as the focal point) for pointing, and eyebrow up and mouth up for selecting the keys for keyboard input. They found no significant difference between mouth and eyebrow selection in performance, however, some participants could not perform the eyebrow up gesture for too long. Our application also utilizes head tracking using nose as the focal point, and eyebrow and mouth gestures. [13] is another similar work which also found the eyebrow-based gesture (eyebrow down in particular) failing to work for most participants. We decided that using eyebrow gesture for the most used actions might not be suitable, and it should only be reserved for the actions performed rarely. In our media player, we use it for volume control.

Likewise, [10] compares facial navigation and selection methods to select the provided targets on the screen. Similar to [12], they move the cursor using head movement, and smile, blink and dwell for selecting a target (similar to left click action in a mouse). They found smile selection working the best and performing better than blink. [19] shows that using blink and wink as input methods is viable with the current technology. They compare replacing the mouse clicks with winks (left wink for left click and right wink for right click), from which they found that sometimes winking at the right moment could be difficult for some people. In [17], they develop a system for generating notes from lecture videos using blink interaction which achieved good usability and accuracy scores.

[20] is yet another work that allowed users to enter text and browse the Internet using gaze and blink interaction. They performed an experiment on healthy as well as disabled people (including people with athetosis). While the main focus of this work was on the technicalities, the positive results highlight the potential of this type of interaction for people with disabilities. In [22], they design a device that aids paralyzed people to communicate using blink. The device detects and converts blinks to Morse code, which is then mapped to English alphabets. It shows that such methods could be highly effective for paralyzed people. This further strengthens the point that such forms of interactions can be highly effective for people with disabilities or for hands-free input.

[7] is a very closely related work to this paper which looks at how to design voice interactions for "how-to" videos. Although they have used voice as the input method, they are also trying to solve navigation problems in videos while the user cannot use their hands. They have listed the commands they used for video navigation. It is similar to the list of commands posted as a guideline in [24]. In our prototype, we map our face-based techniques to commands similar to what they have done with voice in this paper.

We found mouth open, blinks, winks and eyebrows as viable facial gestures to be used in our prototype. In almost all of these papers, the common measurements include accuracy (usually as the number of failures or attempts required), task completion time, fatigue, and the user rating (how they perceive the ease of use, or enjoyment in the case of games). In a majority of these papers, they

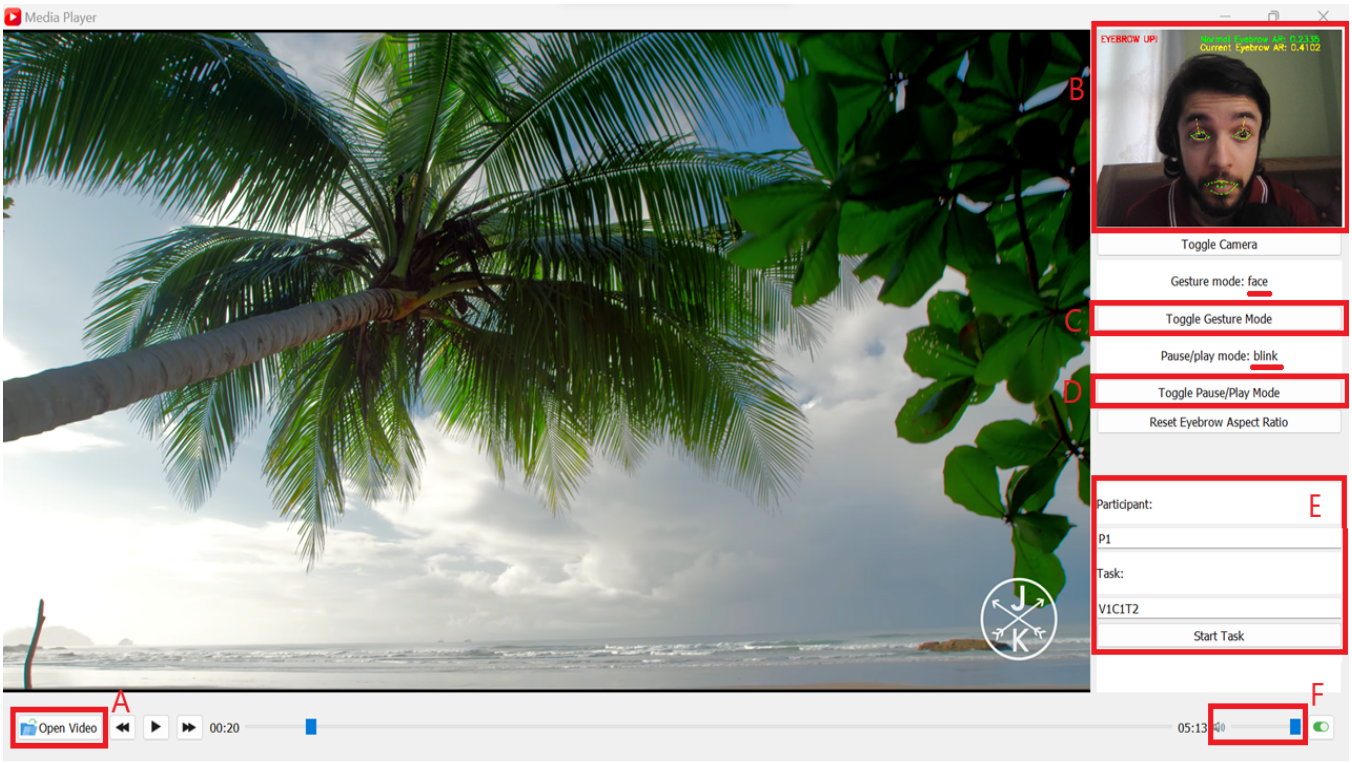

Figure 2: Media Player Interface. A: Button to pick a video from the file system, B: The camera interface, which also shows the current input if detected, C: Button to change the navigation mode between nose and face, D: Button to change the pause/play mode between mouth and blink, E: Area to include participant and task names (used for logging the time taken to complete a task), F: Volume slider which also shows the current volume level

have compared their proposed interaction technique with other existing baseline techniques such as mouse, keyboard, touch and tilt. In our case, we decided to compare different face-based interaction techniques with each other to find the best combination.

3 FACE MEDIA PLAYER

To see which facial features or gestures perform better, we created a media player using PyQt5 [8] for the GUI, dlib library [18] to detect facial landmarks, and OpenCV [3] for image processing. The initial setup was taken from [6] and modified to fit our experiment's needs. The gestures used are mouth, wink, blink, and eyebrow up/down. Additionally, nose position is tracked and when it goes beyond the bounding box, a command is triggered. The commands used are play/pause, volume up/down and forward/backward. Table 1 shows the list of actions and the respective commands they are mapped to.

We used a subset of commands from [7] with the exclusion of video speed control, mute, and stop.

3.1 Configurations

There are two configurations for the pause/play command: in the mouth configuration, opening the mouth triggers pause or play depending on the current state, and in the blink configuration, blinking three times in a row within a second does the same thing.

Table 1: Action to Command Mapping

Action	Command
Mouth Open	Pause / Play
Blink Thrice in a Second	Pause / Play
Eyebrow Up / Down	Volume Up / Down
Nose Up / Down	Volume Up / Down
Wink Left / Right	Backward / Forward
Nose Left / Right	Backward / Forward

We also have two configurations for volume control and forward/backward (we jointly call them navigation): the nose configuration, where moving the nose towards right or left of the bounding box triggers forward or backward command in the media player, and moving the nose up or down triggers volume up or down respectively; and the face gesture configuration where right and left winks trigger the forward and backward action, and eyebrow up and down trigger volume up and down respectively.

3.2 Parameters

Here we describe the parameters we used for the experiments. The forward and backward commands trigger the respective functions in the media player, skipping or rewinding the video by 5 seconds.

To account for longer videos, we added a constant acceleration of 1 for both nose left/right and wink, where each successive forward or backward increased the skip/rewind factor by 5 seconds. The maximum skip/rewind factor was set to 2 minutes. This was reset once the action was stopped. The volume up and down actions increased or decreased the volume by a factor of 10 respectively, with 0 being the lower and 100 being the upper bounds.

Likewise, for the blink play/pause mode, the first blink triggers counting of blinks and the next two successive blinks within a second triggers the pause/play command. Failure to detect blink thrice within a second resets the counter and the user has to start blinking again. Similarly, for the mouth open mode, once the user opens the mouth and the pause/play command is triggered, the application waits for 1 second before triggering the next pause/play command.

With these configurations and parameters, we compare mouth open and blink for the pause/play command, eyebrow and nose up/down actions for volume control, and wink left/right and nose left/right actions for moving forward and backward across the video.

4 USER STUDY

Based on the previous work on wink, blink, and eyebrow-based interactions, we hypothesized that the nose-based navigation techniques and the mouth gesture would outperform their counterparts in terms of the overall task workload, enjoyment, perceived accuracy, and the ease of use. We expected to see no significant difference between these methods in terms of the actual task completion time. For internal validity, we performed a controlled experiment with 12 participants.

4.1 Participants

12 unpaid participants (9 male, 3 female) aged between 20 to 30 years (M=26,SD=2.37) participated in our experiment. 5 of the participants wore eye glasses. The participants were novices in regards to the face-based input methods. All the participants were familiar and frequent users of media players.

4.2 Experimental Setup

Figure 1 shows the experimental setup for our experiments. The experiments were done in a controlled environment with same lighting conditions for each participant. We used a Dell-Inspiron 5558 laptop with an Intel Core i5 and with 8GB of RAM. The display size was 15.6 inches and the display resolution was 1366x768 pixels. Throughout all the experiments, the questionnaire data was recorded on a different laptop.

4.3 Experimental Design

The experiment was a 2x2x2 within subject design. The IVs were video length (10 and 60 minutes), navigation method (nose tracking and face gestures), and pause/play method (mouth open and blink). The experiment comprised two sessions (video length: 10 minute and 60 minutes) and four conditions (Mouth + Nose, Mouth + Face, Blink + Nose, and Blink + Face). The videos were ordered sequentially according to the length. The remaining conditions were counterbalanced using a Balanced Latin Square.

4.4 Tasks and Procedure

Upon arrival, participants filled in the demographic questionnaire where we asked about their gender, age, eye condition (glasses or no glasses), and previous experience with face-based interaction. Following this, they were introduced to the software and were allowed to interact with it for 5 minutes. The participants then started with their first condition. For each condition, the participants were asked to perform two sets of pause/play, two sets each of forward and backward, and three sets of volume up/down commands. The order of tasks are listed in Table 2. For the forward and backward tasks, the participants were allowed to be off by 20 seconds from the actual target time.

Table 2: Task Order

Task Name	Task
T1	Play-Pause
T2	Forward to Third Quarter
T3	Volume Max
T4	Backward to First Quarter
T5	Volume Low
T6	Forward to End
T7	Play-Pause
T8	Backward to Start
T9	Volume Medium

Before each task, the experimenter entered the current task and condition name in the media player. The participants were then instructed to click the start task button before beginning each task and press stop once they were done with the task. After each condition, the participants were asked to fill up a subset of the NASA TLX questionnaire [16] for assessing the task workload. At the end, the participants filled up a usability questionnaire [4] and also answered some semi-structured interview questions.

5 RESULTS

5.1 Task Completion Time

We compared the time to complete the tasks between each experiment condition with a two-way repeated measures ANOVA test. (For each ANOVA we did during this experiment, the normality and sphericity assumptions have been checked, in case of violation of the sphericity assumptions, the Greenhouse-Geisser correction is used.) The test exhibits a significant difference of the completion time among the experimental conditions. $F(1.729, 19.127) = 4.9, p < 0.05, \eta^2 = 0.312$ The strong value of η^2 indicates a strong effect size of this factor among the samples. Regarding the video length, no significant differences has been noted during the experience: $F(1, 11) = 3.492, p > 0.05, \eta^2 = 0.241$. The value of the partial eta squared indicates a strong effect size of the video length too. For the entire population, the condition Conditions*Video of the two-way ANOVA implies no interaction effect was been noted. $F(1.827, 20.094) = 0.96, p > 0.05, \eta^2 = 0.08$.

Then for each command: play (pause), volume up (down), forward (backward) we compare the input mode two by two. That means for play/pause we compare the mouth open gesture with the

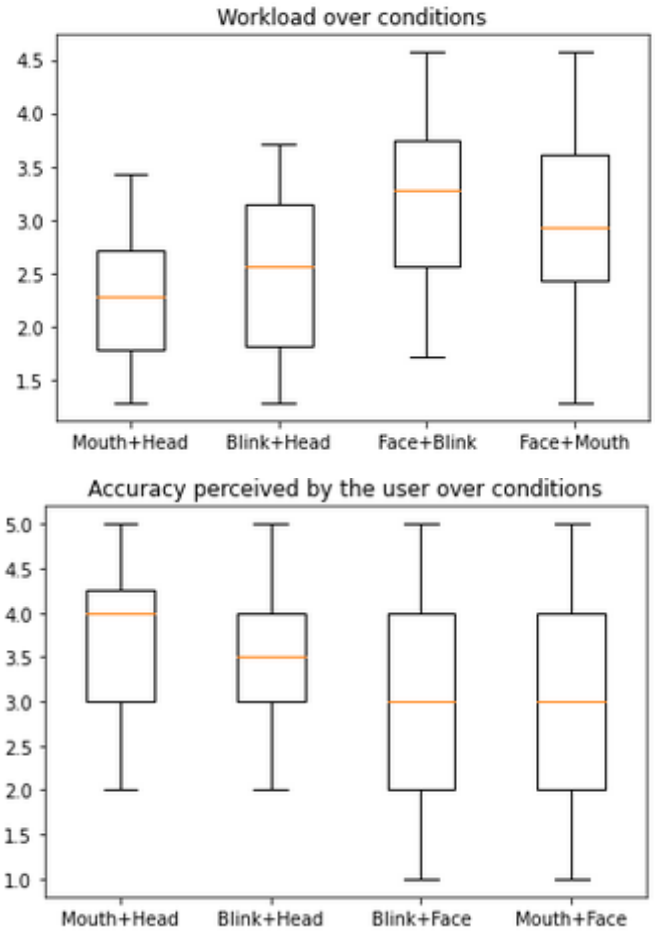

Figure 3: (a) workload (b) accuracy perceived

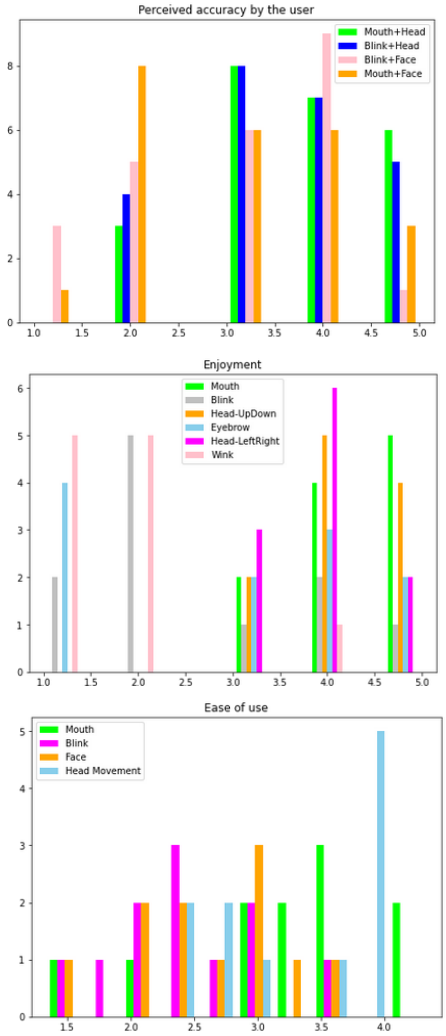

Figure 4: (a) accuracy (b) enjoyment (c) ease of use

three blinks in a row, for volume up/down we compare eyebrow up/down with nose vertical movement and for forward/backward we compare the winking gesture with the nose horizontal movement.

The two-way repeated measures ANOVA for the play/pause command shows no significant differences between the two input mode. $F(1, 11) = 1.306, p > 0.05, \eta^2 = 0.106$. The test on the video length show a significant differences between the two population, $F(1, 11) = 8.961, p < 0.05, \eta^2 = 0.449$.

The two-way repeated measures ANOVA for the volume up/down command shows no significant differences between the two input mode. $F(1, 11) = 0.030, p > 0.05, \eta^2 = 0.003$. The test on the video length show a significant differences between the two population, $F(1, 11) = 6.419, p < 0.05, \eta^2 = 0.368$.

The two-way repeated measures ANOVA for the forward/backward command shows a significant difference between the two input mode. $F(1, 11) = 9.593, p < 0.05, \eta^2 = 0.466$. The test on the video length show no significant differences between the two population, $F(1, 11) = 0.706, p > 0.05, \eta^2 = 0.060$.

5.2 Questionnaire Feedback

The results from the questionnaire are presented in Figures 3, 4, and 5. For each command for each input mode and each category of feedback we performed a paired sample T-Test with the following alternative hypothesis (H_1: $x_1 \neq x_2$ to check if there is significant difference or not. For the play/pause command the paired sample T-Test shows a significant difference, in terms of workload $t(47) = -3.691, p < 0.5$, also in terms of enjoyment $t(10) = 5.19, p < 0.05$, and in terms of ease-of-use $t(10) = 3.088, p < 0.05$. For the forward/backward command, a significant difference was found in terms of workload $t(47) = 6.049, p < 0.05$, of enjoyment $t(10) = 5.449, p < 0.05$ and in terms of ease-of-use, $t(10) = -3.055, p < 0.05$. For the volume up/down, only a significant difference have been found in terms of enjoyment $t(10) = 2.609, p < 0.05$.

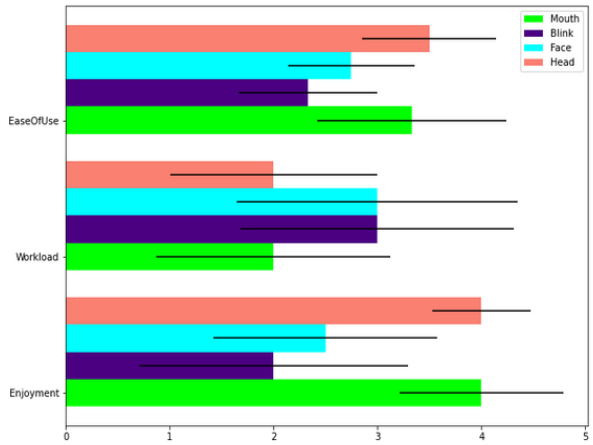

Figure 5: Summarization of the feedback over the different type of input

6 DISCUSSION

6.1 Task Completion Time

From our results, we found that there was a significant difference in task completion time for the forward and backward tasks between wink and nose left/right. Wink was much slower in contrast to what we had believed before the experiments. However, some participants were able to complete the tasks using wink much faster than others, also beating the performance of nose. We believe this is because it could be difficult for some people to wink than others. One might also speculate that a person's facial structure would make it either easier or more difficult to detect gestures such as wink, blink, and eyebrow movements. Whereas the nose tracking is unaffected, and, in theory, should perform equally well for any individual, regardless of their facial structure.

The remaining results went according to our hypotheses. While the average task completion time for play/pause using mouth open was faster than three blinks in a row, the difference was not statistically significant. Similarly, there was no difference between eyebrow up/down and nose up/down for volume control as even the averages were very close to each other. An unexpected results is the significant difference for the video factor during the experiment on the play (pause) and volume up (down) commands. Indeed, during the designing we create the lengths factor to see how it will impact the forward (backward) operation. The significant differences got are certainly due to a learning effect across the experiment of these two tasks.

6.2 Qualitative Feedback

In terms of the task workload, which we assessed through questionnaires after each condition, we found the nose movement and mouth open gesture outperforming their counterparts significantly. The participants felt that there was a higher mental as well as physical demand while using the facial gestures instead of nose movement, and blink instead of mouth open. The best combination was Mouth + Nose movement, with the worst combination being Blink + Face. The other metrics that we assessed through the questionnaires also show a similar trend, with mouth and nose movement being the clear favorites in terms of ease of use, enjoyment, and perceived accuracy. This stays true to our hypotheses.

Through the interviews, we learned that most participants preferred the nose tracking method because they felt like they did not have to focus on accurately producing a gesture, and instead could just focus on the task. Whereas for the wink, they had to be more aware about their head position and sometimes required multiple attempts for the wink to work as intended. All the participants mentioned that wink was more difficult to perform than nose movement. The answers were similar when comparing mouth open to blink, with mouth open being the preferred choice for every participant. About half of the participants felt that it was equally easy to perform the eyebrow gestures compared to nose up/down for volume control, with three of them even preferring the eyebrow volume control over nose up/down. They felt like it required less effort than moving the head up and down.

6.3 Limitations and Future Work

We noticed that the average time to complete the tasks, including the forward/backward was much lower for the longer video than the shorter video even though the user had to cover a longer duration. This might be due to the learning effect, as the longer video always came after the shorter one, and also because the acceleration (which we talked about in Section 3) increased the sensitivity, making it difficult to navigate the shorter video. Two of the participants had also mentioned in the interviews that they felt like the timer was moving too quickly as they tried to reach a particular point in time. We believe this to be one of the limitations of our work. Coming up with a perfect acceleration algorithm, which works for videos of variable lengths might be one possible direction of work which would improve the user experience and could be one of the steps to make the techniques we discussed in this paper viable in uncontrolled environments for media players or other similar applications.

Another thing we have to consider is that the participant feedback could be biased because of how these techniques are implemented. If we were to solely compare these techniques in terms

of participant feedback, it would perhaps be better to conduct a Wizard of Oz study where the media player is controlled behind the scenes by a real human without the user's knowledge. We consider it to be one of the limitations of our work as, if the time permits, it might be better to split the experiments into two different studies, where the first study compares the participant feedback and the second assesses the quantitative data such as task completion time and accuracy.

As we discussed in sections 1 and 2, the users who might benefit the most from this kind of work are disabled people who are unable to use their hands. It might be appropriate to perform a controlled experiment with disabled people to assess their performance and feedback.

Finally, we have to point out that in this experiment, we have compared nose tracking to wink and eyebrows. Unlike other papers that have mostly used nose/head tracking for cursor movement, we have a one-to-one mapping between the nose horizontal and vertical movement and facial gestures to control the media player. Nonetheless, one might argue that a closer technique to compare nose tracking to is something like gaze tracking. Now that we know nose clearly outperforms facial gestures, one possible direction of work might be comparing it to using gaze tracking for media player control.

6.4 Suggestions

In order to design a face-based media player, we suggest the use of nose movement for navigating through the video and volume control, with the possibility of using eyebrow gestures instead for volume control. We suggest the use of mouth open for play/pause command. Additionally, for other features, such as increasing or decreasing the video speed or stop command, it might be possible to use other gestures we discussed in this paper assuming that these tasks are not as frequently used by the users.

7 CONCLUSION

Quoted from [12], video-based interaction, particularly, face-based interaction is indeed one promising technology that supports hands-free interaction. In this paper, we looked at how such face-based interaction techniques compare against each other in a niche use case like controlling a media player. We looked at which technique is better in terms of task completion time, task workload, ease of use, enjoyment, and perceived accuracy.

We created a media player with different face-based interaction techniques mapped to the commands in the media player. We conducted a controlled experiment with 12 participants to see how these different face-based techniques perform. We used mouth and blink for pause/play, nose horizontal movement and wink for forward/backward, and nose vertical movement and eyebrow movement for volume control. We also tested different combinations of these techniques with two videos of different lengths.

Nose horizontal movement had a lower task completion time than wink, with other techniques performing similarly. Throughout all our assessments of the user feedback, we found nose movement and mouth open gesture outperforming their counterparts in wink, eyebrow, and blink. Participants mentioned the wink and blink in particular was more difficult to perform and indeed fatiguing, which is consistent to what we saw in previous work. We conclude that Nose movement + Mouth open gesture is the best combination to control the media player.

Similar to previous work, our work strengthens the point that gestures such as wink might be difficult to perform. We expect these methods to work better naturally with further improvements in Computer Vision techniques. In the future, we plan to observe similar techniques applied to other similar use cases such as a PDF reader. From this experience, we also find it appropriate to increase the range of techniques used for comparison as there might be other suitable gestures that perform better for more individuals than blinks and winks. The most obvious continuation of this work would be to explore more face-based interaction methods in order to find the perfect combination for controlling a media player.

REFERENCES

[1] Mahdieh Abbaszadegan, Sohrab YaghoubiI, and I. Scott MacKenzie. 2018. TrackMaze: A Comparison of Head-Tracking, Eye-Tracking, and Tilt as Input Methods for Mobile Games. Kurosu M. (eds) Human-Computer Interaction. Interaction Technologies. In *HCI 2018. Lecture Notes in Computer Science, vol 10903.* 393–405. https://doi.org/10.1007/978-3-319-91250-9_31

[2] M. Betke, J. Gips, and P. Fleming. 2002. The Camera Mouse: visual tracking of body features to provide computer access for people with severe disabilities. *in IEEE Transactions on Neural Systems and Rehabilitation Engineering* 10, 1 (March 2002), 1–10. https://doi.org/10.1109/TNSRE.2002.1021581

[3] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[4] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).

[5] Dario Cazzato, Marco Leo, Cosimo Distante, and Holger Voos. 2020. When I Look into Your Eyes: A Survey on Computer Vision Contributions for Human Gaze Estimation and Tracking. *Sensors* 20 (2020), 13. https://doi.org/10.3390/s20133739

[6] Akshay L Chandra and Harshitaracha. 2022. Mouse Cursor Control Using Facial Movements. https://github.com/acl21/Mouse_Cursor_Control_Handsfree.

[7] Minsuk Chang, Anh Truong, Oliver Wang, Maneesh Agrawala, and Juho Kim. 2019. How to Design Voice Based Navigation for How-To Videos. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, , NY, USA, Paper 701.* 1–11.

[8] Riverbank Computing. 2021. PyQt5. https://pypi.org/project/PyQt5/

[9] Justin Cuaresma and I. Scott MacKenzie. 2014. A comparison between tilt-input and facial tracking as input methods for mobile games. 2014 IEEE Games Media Entertainment. *Toronto, ON, Canada* (2014), 1–7. https://doi.org/10.1109/GEM.2014.7048080

[10] J. Cuaresma and I. S. MacKenzie. 2017. FittsFace: Exploring Navigation and Selection Methods for Facial Tracking. In *Antona M*, C. Stephanidis (Ed.). Universal Access in Human–Computer Interaction. Designing Novel Interactions. UAHCI 2017. Lecture Notes in Computer Science, vol 10278. Springer, Cham. https://doi.org/10.1007/978-3-319-58703-5_30

[11] Debjyoti Ghosh, Can Liu, Shengdong Zhao, and Kotaro Hara. 2020. Commanding and Re-Dictation: Developing Eyes-Free Voice-Based Interaction for Editing Dictated Text. ACM Trans. *Comput.-Hum. Interact* 27 (August 2020), 4. https://doi.org/10.1145/3390889

[12] Yulia Gizatdinova and Oleg Špakov Veikko Surakka. 2012. Comparison of video-based pointing and selection techniques for hands-free text entry. In Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI).

[13] Yulia Gizatdinova and Oleg Špakov Veikko Surakka. 2014. Face typing: Vision-based perceptual interface for hands-free text entry with a scrollable virtual keyboard.

[14] Argenis Ramirez Gomez and Hans Gellersen. 2020. More than Looking: Using Eye Movements Behind the Eyelids as a New Game Mechanic. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play.* Association for Computing Machinery, New York, NY, USA, 362–373.

[15] Kristen Grauman et al. 2003. Communication via eye blinks and eyebrow raises: video-based human-computer interfaces. *Universal Access in the Information Society* 2 (2003), 359–373. https://doi.org/10.1007/s10209-003-0062-x

[16] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[17] P. Kar, S. Banerjee, S. Chakraborty, et al. 2021. AutoNotes: A Touch-Free Blink-Based Interactive Model for Generation of Notes from Lecture Videos. *J. Inst* 102

(2021), 1157–1166. https://doi.org/10.1007/s40031-021-00550-4
[18] Davis King. 2022. dlib. https://pypi.org/project/dlib/
[19] Piotr Kowalczyk and Dariusz Sawicki. 2019. Blink and wink detection as a control tool in multimodal interaction. *Multimedia Tools Appl.* 78, 10 (May 2019), 2019. https://doi.org/10.1007/s11042-018-6554-8
[20] A. Królak and P. Strumiłło. 2012. Eye-blink detection system for human–computer interaction. *Univ Access Inf Soc* 11 (2012), 409–419. https://doi.org/10.1007/s10209-011-0256-6
[21] Pedro Monteiro, Guilherme Gonçalves, Hugo Coelho, Miguel Melo, and Maximino Bessa. 2021. Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2702–2713. https://doi.org/10.1109/TVCG.2021.3067687
[22] K. Mukherjee and D. Chatterjee. 2015. *Augmentative and Alternative Communication device based on eye-blink detection and conversion to Morse-code to aid paralyzed individuals. 2015 International Conference on Communication.* Information Computing Technology (ICCICT). . https://doi.org/10.1109/iccict.2015.7045754
[23] Munir Oudah, Ali Al-Naji, and Javaan Chahl. 2020. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. *Journal of Imaging* 6 (2020), 8. https://doi.org/10.3390/jimaging6080073
[24] Javier Pérez. 2022. Voice commands guide. https://www.handsfreeforweb.com/en/commands-guide.html
[25] Sean-Kerawala and TimShererWithAquent. 2022. Hands-free. https://docs.microsoft.com/en-us/windows/mixed-reality/design/hands-free
[26] Terry Stancheva. 2022. 24 Noteworthy Video Consumption Statistics [2022 Edition]. https://techjury.net/blog/video-consumption-statistics
[27] Eduardo Velloso, Amy Fleming, Jason Alexander, and Hans Gellersen. 2015. Gaze-Supported Gaming: MAGIC Techniques for First Person Shooters. In *15). Association for Computing Machinery, New York, NY, USA.* Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY, 343–347.
[28] Bryan Wang and Tovi Grossman. 2020. BlyncSync: Enabling Multimodal Smartwatch Gestures with Synchronous Touch and Blink. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–14.
[29] Wikipedia. 2022. Hands-free computing — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=Hands-free%20computing&oldid=1000081301. [Online; accessed 26-March-2022].