Facial Blendshapes for Custom Action: An Experimental Evaluation Using Facial Expressions for Human-Computer Interaction

https://doi.org/10.31713/MCIT.2024.051

Serhii Zelinskyi
Department of Computer Engineering
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
serhii.zelinskyi@knu.ua

Yuriy Boyko
Department of Computer Engineering
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
boyko@univ.kiev.ua

***Abstract* — This study presents a novel human-computer interaction approach that uses facial blendshapes and facial expressions to define custom actions. By using predefined facial expressions and their combinations, users can trigger custom actions in a simulated text editor. This approach can provide an accessible interaction method for individuals with upper limb disabilities and potentially offer productivity benefits for regular users. The suggested approach was implemented and evaluated by conducting an experiment involving six participants. The study compared task completion time using traditional mouse input and a combination of mouse and facial expressions. Preliminary results showed that while using facial expressions requires more time to master, participants valued the hands-free interaction potential of this approach.**

Keywords — **human-computer interaction; facial blendshapes; facial expressions; accessibility; hands-free control.**

I. Introduction

The need for more productive and accessible ways to interact with digital systems has made researchers look for new methods beyond traditional devices like keyboards and computer mice. For people with upper limb disabilities, it can be challenging to use regular input devices. This motivates finding new solutions that allow them to interact with digital systems without their hands. At the same time, regular users are looking for ways to increase their productivity, for example, by using shortcuts for tasks they frequently do.

While gesture and voice recognition systems are popular interaction methods, facial expressions offer a new, underused way to control digital systems without using hands. This study suggests using facial blendshapes [1], which recognize facial movements based on key points of the face, to define custom action. This method can give individuals with upper limb disabilities another way to interact with digital systems and help increase productivity for regular users while completing tasks.

II. Related Work

Facial expressions have become a promising modality for human-computer interaction, offering new ways of interaction with digital systems. Recent studies have explored various approaches to capturing and interpreting facial expressions. In [2], a system using electromyography to detect muscle activity and control a virtual avatar was developed. Similarly, smart glasses equipped with optomyographic sensors were introduced to recognize facial gestures for device interaction [3]. Several studies have developed systems to simulate computer mouse behavior using face gestures [4], [5], [6]. Some of these systems combined face gestures with gaze pointing [7], while others integrated speech-to-text functionality [8].

However, systems based on predefined facial gestures are limited by the fixed number of actions they can recognize, reducing flexibility in interaction. On the other hand, facial blendshapes capture subtle variations in facial expressions, allowing for a broader space of actions to be defined.

III. Methodology

The suggested approach has been implemented in the form of a web application. It utilizes MediaPipe Face Landmarker [9] to detect blendshape scores, which represent detailed facial deformations. These blendshapes were used to define custom facial expressions that mapped to actions in a simulated text editor. Each action is represented by either a single facial expression or a combination of two facial expressions (indicated by the “+” sign). Facial expressions and their corresponding actions are listed below:

- **Left Eye Closed** — apply the “Bold” style to the selected words.
- **Mouth Pucker** — apply the “Italic” style to the selected words.
- **Smile** — apply the “Underscore” style to the selected words.
- **Jaw Open** — apply the “Strikethrough” style to the selected words.
- **Left Eye Closed + Smile** — apply the “Highlight” style to the selected words.
- **Left Eye Closed + Mouth Pucker** — delete the selected words.
- **Left Eye Closed + Jaw Open** — select all words in the text.

Fig. 2 contains a screenshot of the experimental setup user interface.

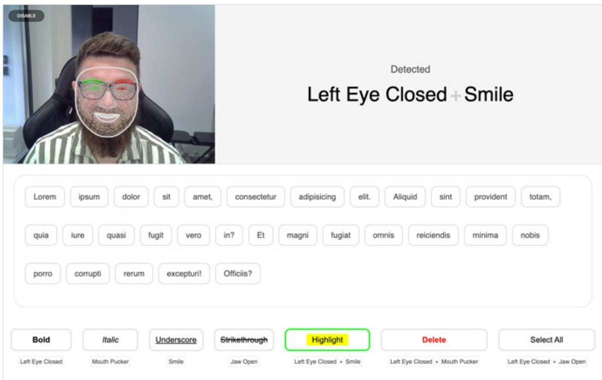

Figure 1. Screenshot of the simulated text editor interface used in the experimental setup

In the experimental setup, users could select individual words using a mouse and apply formatting, text modification, or selection actions by clicking on the corresponding buttons or using custom facial expressions.

To make applying actions using facial expressions more reliable, we designed an interaction method that requires users to hold down the "Shift" key on their keyboard while making and maintaining the facial expression. The system displays the facial expressions it recognizes in real-time and highlights the corresponding action button in the user interface. When users are confident that the system has correctly recognized the action they want to apply, they release the "Shift" key, and the corresponding action is applied to the selected words. This additional step ensures accuracy and gives users control over the action before it is executed.

This approach was implemented in the current experiment for evaluation purposes. To make the system fully hands-free, future implementations could replace the "Shift" key with a voice command or other input modalities, such as a blink or nod gesture. This would allow users, particularly those with upper limb disabilities, to interact with the system more naturally and enhance the system's accessibility.

The blendshape scores provided by the Face Landmarker allow for real-time detection of facial movements, triggering the appropriate actions based on the defined facial expressions. This interaction method provides users with a hands-free alternative to traditional input methods while maintaining a level of control and reliability in applying the desired actions.

IV. Evaluation and Results

An experiment was conducted with six participants, all of whom were regular users with no upper limb disabilities. The purpose of the experiment was to validate the suggested approach by evaluating its effectiveness in applying text formatting and text manipulation actions and comparing the results of using mouse-only input versus a combination of mouse and facial expressions input. Each participant was tasked with performing a sequence of actions, explained in the Methodology section, using both input methods. Task completion times were recorded for each participant across both input methods. In addition, participants were asked to provide feedback on the usability and intuitiveness of the facial expression system.

The conducted experiment demonstrated the following results:

- The average task completion time using mouse-only input was **20.34 seconds**.
- The average task completion time using mouse and facial expressions was **33.45 seconds**.

Key points from the participant feedback included the following:

- **Memorization difficulty.** Most participants mentioned that they had difficulties in remembering what facial expression corresponded to each action, particularly when using combined facial expressions (e.g., Left eye closed + Smile for "Highlight"). So, they had to rely on visual hints to complete tasks accurately.
- **Suggestions for improvements.** Participants suggested enhancing the system by providing more intuitive mappings between facial expressions and actions or displaying better visual cues.
- **Hands-free interaction benefits.** Participants valued the "hands-free" interaction as an alternative to traditional input methods, especially in multitasking or when physical limitations are a factor.

Fig. 2 contains a bar chart illustrating task completion times for each participant across both input methods.

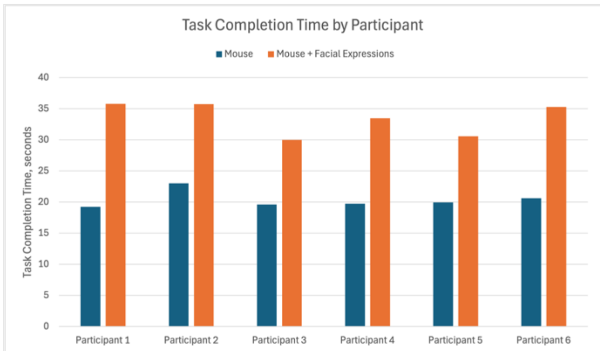

Figure 2. Task Completion Time for Mouse vs. Mouse + Facial Expressions Across Participants

V. Discussion

This study demonstrated both strengths and challenges of using facial expressions with blendshapes for hands-free interaction. While the approach works and has clear benefits, the conducted experiment highlighted areas that need improvements.

Advantages. The main advantage of the system is that it offers hands-free control, which is especially useful for people with upper limb disabilities or when multitasking. By using facial blendshapes, the system is capable of recognizing small facial movements, which allows the creation of a variety of custom actions by using blendshapes related to eyes, brows, cheeks, mouth, jaw, and their combinations. This flexibility allows users to create shortcuts, which can increase their productivity.

The use of a standard web camera makes the system more accessible and low-cost.

Disadvantages. The primary limitation is the learning curve, as participants found it difficult to memorize facial expressions, leading to slower task completion times. The performance of the system can be affected by its sensitivity to light conditions or by using a low-quality web camera. The need to hold the "Shift" key when applying an action using facial expressions increases the system's reliability but also adds complexity, which results in increased task completion time.

Potential improvements. The usability of the system can be improved with more intuitive facial expression mappings and improved real-time visual feedback when applying actions. An improved process of applying facial expression actions, instead of holding the "Shift" key, could reduce task completion time, enhancing the user's overall performance.

Application. The suggested approach can offer a new way of interacting with digital systems for people with upper limb disabilities. It can also be helpful in situations where using a computer mouse or keyboard is limited, such as presentations or in industrial settings. The ability to create custom shortcuts could improve productivity in many different areas.

VI. Conclusion and Future Work

This study introduced a new way of human-computer interaction using facial blendshapes and expressions. The conducted experiment involving six participants demonstrates that while the system offers a useful hands-free interaction method and flexibility through custom facial expressions, there is a learning curve that slows down task completion, as participants took more time to complete tasks with facial expressions input than with the mouse only.

However, the participants saw the system's potential to increase productivity, especially in situations when traditional input methods, like the keyboard or mouse, might not be practical. The ability to combine facial expressions allows for a variety of custom actions. That makes the system flexible for different user needs, including individuals with upper limb disabilities.

Future research could explore using a facial expression sequence to define custom actions, which can significantly expand the space of possible custom actions. Also, combining facial expressions with other input modalities, like voice, gestures, head pose, body pose, or gaze (for example, for pointing), could make the system even more flexible and improve the user experience.

References

[1] J. P. Lewis, K. Anjyo, T. Rhee, M. Zhang, F. Pighin, and Z. Deng, 'Practice and Theory of Blendshape Facial Models', 2014, *The Eurographics Association*. doi: 10.2312/EGST.20141042.

[2] F. French *et al.*, 'Expressive Interaction Design Using Facial Muscles as Controllers', *Multimodal Technol. Interact.*, vol. 6, no. 9, p. 78, Sep. 2022, doi: 10.3390/mti6090078.

[3] H. Gjoreski *et al.*, 'OCOsense Glasses – Monitoring Facial Gestures and Expressions for Augmented Human-Computer Interaction: OCOsense Glasses for Monitoring Facial Gestures and Expressions', *Ext. Abstr. 2023 CHI Conf. Hum. Factors Comput. Syst.*, pp. 1–4, Apr. 2023, doi: 10.1145/3544549.3583918.

[4] H. Mosquera, H. Loaiza, S. Nope, and A. Restrepo, 'Identifying facial gestures to emulate a mouse: Control application in a web browser', *2016 XXI Symp. Signal Process. Images Artif. Vis. STSIVA*, pp. 1–6, Aug. 2016, doi: 10.1109/STSIVA.2016.7743345.

[5] A. Dongre, R. Pinto, A. Patkar, and M. Lopes, 'Computer Cursor Control Using Eye and Face Gestures', *2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT*, pp. 1–6, Jul. 2020, doi: 10.1109/ICCCNT49239.2020.9225311.

[6] A. Kumar Raja, C. Sugandhi, G. Nymish, N. Sai Havish, and M. Rashmi, 'Face Gesture Based Virtual Mouse Using Mediapipe', *2023 IEEE 8th Int. Conf. Converg. Technol. I2CT*, pp. 1–6, Apr. 2023, doi: 10.1109/I2CT57861.2023.10126453.

[7] D. Rozado, J. Niu, and M. Lochner, 'Fast Human-Computer Interaction by Combining Gaze Pointing and Face Gestures', *ACM Trans. Access. Comput.*, vol. 10, no. 3, pp. 1–18, Aug. 2017, doi: 10.1145/3075301.

[8] L. R. Kalabarige, K. A. Abhilash, K. A. Trivedi, and M. Dathatreya, 'Facial Landmark-based Cursor Control and Speechto-Text System for Paralyzed Individuals', *2023 Int. Conf. Sustain. Comput. Data Commun. Syst. ICSCDS*, pp. 849–856, Mar. 2023, doi: 10.1109/ICSCDS56580.2023.10104936.

[9] 'Face landmark detection guide', Google AI for Developers. Accessed: Sep. 06, 2024. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker