Defining a mid-air gesture dictionary for web-based interaction

Thomas Pasquale
University of Turin
Turin, Italy
thomas.pasquale@unito.it

Cristina Gena
Computer Science Department,
University of Turin
Turin, Italy
cristina.gena@unito.it

Fabiana Vernero
Computer Science Department,
University of Turin
Turin, Italy
fabiana.vernero@unito.it

ABSTRACT

This paper presents an empirical evaluation of mid-air gestures in a web setting. Fifty-six (56) HCI students were divided into 16 groups and involved as designers. Then, they proposed a set of mid-air gestures to carry out the identified actions: 99 different mid-air gestures for 16 different web actions were produced in total. Designers validated their proposals involving external subjects, namely 248 users in total. Finally, we analyzed their results and identified the most recurring or intuitive gestures as well as the potential criticalities associated with their proposals.

CCS CONCEPTS

• **Human-centered computing** → **Gestural input**.

KEYWORDS

touchless interactions, gesture recognition, gesture dictionary

ACM Reference Format:
Thomas Pasquale, Cristina Gena, and Fabiana Vernero. 2024. Defining a mid-air gesture dictionary for web-based interaction. In *International Conference on Advanced Visual Interfaces 2024 (AVI 2024), June 03–07, 2024, Arenzano, Genoa, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3656650.3656661

1 INTRODUCTION

Touchless interaction is becoming a popular way to interact with devices, with no need for physical contact. This is especially true in public spaces, where touchless interaction can be faster than using a mouse or a keyboard, also due to the frequent use of large displays [2]. According to Ardito et al. [2], hand gestures are those whose interaction modality is through remote gestures (namely performed by the user without any contact with the display, thus touchless), and they are also called mid-air, and so we will refer to them.

Web browsing is one of the most common tasks that users perform on their devices. However, touchless web browsing can be challenging, as it can be difficult to move a cursor or carry out other tasks without traditional input devices. Mid-air gestures call indeed for natural interfaces [9], which refer to user interfaces that are invisible, or become invisible with successive learned interactions, to their users, and have been applied in several different domains. While the focus in this paper is primarily on web-related tasks, other studies address the alphabet's representation through gestures, without using an on-screen keyboard [11]. Touchless interaction can help users with disabilities in interacting with devices, for example by translating sign language into text [11] and can also be used in shops to attract customers through innovative forms of interaction and advertising [5]. The use of large displays and mid-air gestures has also been proposed in the context of the Smart Industry [7]. In [1] the project consortium engineered a smart armband able to detect gestures through the analysis of both movement and muscle biosignals, while a machine learning library allows to recognize task-specific gestures.

To interact with a touchless device, gestures must be identified and translated into commands understandable by the system. The use of coloured caps worn on fingertips has been proposed to make the identification of gestures corresponding to cursor actions easier [12]. In this scenario, the left click is produced by moving the index and middle finger close to each other, while the index finger (i.e., with the detection of a single-color cap) is considered enough to perform a right click [12]. Other studies [9] propose the use of the index and middle finger for the double-click confirmation, or a pinch using the thumb and index fingers for drag and drop tasks, while, to control the cursor, the index finger movement is mapped to the representation of the cursor.

System activation can be accomplished in different ways, such as holding the palm open, using a closed fist, making a peace sign, transitioning from a fist to an open palm, or a combination of finger gestures [10]. Instead of proposing a wide range of actions to the user, one way is to limit the actions to elementary ones, such as using the palm for stopping and a thumbs-up gesture for confirmation. Then, the user interface (UI) guides the user through successive choices, further simplifying the interaction but limiting the user's options, as in [3]. Instead of deploying an exhaustive on-screen keyboard interface for textual input, an alternative modality entails the presentation of a miniature virtual keyboard, preconfigured with indispensable typing keys. This pre-mapped keyboard is located on the top of the screen in a red box. Users engage with the system by horizontally sweeping their open palm across the preconfigured keyboard, thereby accessing the alphabet characters and keyboard functions. To select a specific key, users place their finger over the chosen key [8].

This paper presents a large study of mid-air gestures for touchless interaction with a web application. As output, we devised a final dictionary which includes the most suitable mid-air gestures for a fluid web interaction.

AVI 2024, June 03–07, 2024, Arenzano, Genoa, Italy

ACM ISBN 979-8-4007-1764-2/24/06
https://doi.org/10.1145/3656650.3656661

2 THE STUDY

The goal of the study was to define a mid-air gesture dictionary for a web-based interaction, tested in the context of a university classroom search service. Thus, we analyzed a large set of mid-air gestures proposed and validated by a large sample of users, to identify the most prevalent, intuitive, and shared ones.

The study originated in a university setting, during an assignment for a Human-Computer Interaction (HCI) master course. Users were involved on two occasions: firstly, students acted as designers to propose an initial set of mid-air gestures and, secondly, external users, not involved in the design phase, were engaged to collect external feedback. In addition, a few groups also involved external users as co-designers in the design phase. External users were asked to submit a consent form to take part in the research, while students as designers were asked permission to use their material in this research.

Participants. Fifty-six (56) students (71% females) were involved as designers. They were divided into 16 groups (2 to 4 members). All of them had at least a bachelor's degree. Five (5) out of 16 groups (31%) consulted 25 external co-designers (5 per group) to get help in finalizing their proposal. All 16 groups submitted their proposed gestures for external review. The total number of external users involved in this second phase was 248, with a modal value of 4 external users per group, with ages ranging from 18 to 55 years old, 68% females. Their background varies from high school to a bachelor's degree.

Apparatus and Materials. The groups of designers used personal devices, such as laptops, to simulate touchless devices to be used during the mid-air gesture generation phase. Videos showing the designed gestures, to be used to collect external feedback, were captured using an external camera. A web form embedding the videos was used to collect structured feedback about the designed gestures.

Procedure. The study included two macro-phases: the generation of many mid-air gestures for web-based interaction and their evaluation, and the definition of a dictionary optimized for web-based interaction. In the first phase, the following activities were carried out: identification of meaningful web actions, design of mid-air gestures and gesture validation. In the second phase, the main steps were the analysis of the identified actions and the analysis of the proposed gestures for each action.

Mid-air gesture generation. 16 groups of designers were required to design a touchless interaction assuming that a large display was installed on the university campus displaying the web-based classroom schedules.

As a first step, each group had to identify the main actions required to carry out basic tasks in the scenario, as: moving the cursor, clicking confirmation and page resizing. After that, each group had to come up with mid-air gestures that would enable users to carry out such actions. Groups were suggested to reach out to external users for help and inspiration. Thus, 5 groups (31%) chose to consult 25 external co-designers (5 per group), which were asked to enact the gestures they would make on a touchless screen to perform the previously identified actions, and the most popular ones were chosen. The other groups worked autonomously in this phase. Once each group had defined an initial set of gestures, they were required to have them reviewed by external users, to ensure they were easy to understand. To do so, a Google Form was used which included video representations of each gesture to evaluate. More specifically, external users had to associate each gesture with the most appropriate action, choosing from a list which included all the previously identified actions. External users could also provide suggestions for improving gesture execution, either by providing free text comments or by submitting anonymous video proposals. Based on this feedback, the groups had the opportunity to tweak the initially defined gestures if necessary: in particular, 8 groups (50%) further modified their gestures to develop a definitive version.

Finally, each group was required to submit their proposal with a video illustrating all the gestures, and a descriptive report.

Dictionary definition. Firstly, we analyzed the lists of actions proposed by each group, to identify commonalities and merge similar actions. Secondly, we separately considered each action and examined all the devised mid-air gestures, trying to identify commonalities in the proposals of different groups. In both cases, we used categorization and counting [4], an approach to the transposition of quantitative data into qualitative data which implies the definition of recurring categories and can be considered a simplified version of thematic analysis [6]. Then, we closely examined the resulting categories, not only considering their popularity but also trying to anticipate the experience of users performing the defined gestures. Finally, we selected the most suitable and intuitive mid-air gestures for each identified web action, thus defining a dictionary optimized for web-based interaction.

2.1 Results

The designers' groups proposed a set of mid-air gestures including cursor-pointing, click confirmation, scrolling, page resizing, quick history navigation, panning in resized pages (zoom >100%), drag and drop, interaction activation and interaction ending, homepage access, page reload, close current page, context menu access, volume control, multiple items selection, and text selection. In total, 99 different versions of these gestures were designed and analyzed.

However, not all gestures were included in the final dictionary as they were considered unnecessary for achieving optimal and fast interaction. In the following, we will present the gestures we considered for our proposed dictionary.

Cursor pointing. Out of the 16 groups, the majority, comprising 9 groups (56%), opted for the index finger as the virtual representation of the cursor (see Fig. 1). Meanwhile, 4 groups (25%) chose an open hand positioned vertically, synchronizing its movement with the cursor. 2 groups (13%) employed both the index and middle fingers for cursor movement, while one group (6%) utilized the index, middle, and thumb collectively to control the cursor. Although a valid alternative is the use of an open hand, a preference is expressed for the use of the index finger, primarily because of its widespread use in conventional mouse interactions.

Click confirmation. 8 groups out of 16 (50%) choose a quick movement of the index finger toward the webcam. 3 groups (19%) used a double-tap with the index finger in front of the display, while 2 other groups (13%) proposed the same gesture but with two fingers. 2 groups (13%) opted for closing the entire hand over the desired element. 1 group (6%) used the index, middle, and thumb

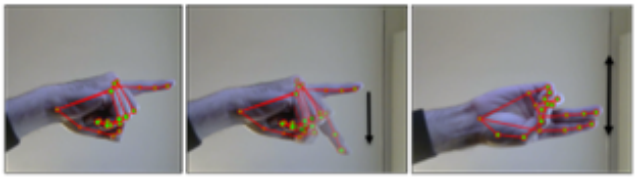

Figure 1: Cursor-pointing, Confirmation, Scrolling. Landmarks were highlighted using Google MediaPipe. (https://developers.google.com/mediapipe/solutions/vision/hand_landmarker).

joining together to complete the action. Building upon the pointing gesture, our investigation was focused on determining the optimal combination to trigger the confirmation action. Given the potential deployment of the system on a kiosk, it is reasonable to assume that end users may not be as close as in the analyzed case but at a greater distance. This adjustment in user positioning should be considered when designing and evaluating the system for practical application at a greater distance. The gesture most frequently observed entails a rapid movement of the index finger towards the display. Although a valid choice, users may adopt an outstretched arm for pointing, and executing the gesture might be impractical without physically moving toward the device. Consequently, this gesture has been excluded from the final dictionary, considering practical concerns. A more suitable alternative could be to employ the second most recurring gesture, namely, an index-finger double tap. Enhancements can be made to this gesture by introducing a timing element for lowering the finger, thereby eliminating the need for a double tap. As depicted in Fig. 1, this gesture is easily replicable in a touchless environment. Nevertheless, it is recommended to augment the range of the index finger movement for a more distinct and easily detectable gesture. Additionally, a minimum holding time of $t >= 2s$ is introduced to execute the action.

Vertical and horizontal scrolling. One group (6%) did not create a gesture for this task. 8 out of the remaining 15 groups (54%) opted to use an open hand perpendicular to the display intending to move vertically or horizontally according to the desired direction. Among these, 3 groups (20%) also design an acceleration in scrolling if the hand movement is intensified. In contrast, all others (80%) presumed a movement directly proportional to the hand's motion. 3 groups (20%) chose to use index and middle fingers in a fixed horizontal position parallel to the webcam, moving in the desired direction. 2 groups (13%) considered a synchronized and repeated movement of pinky, ring, middle and index finger fixed together in the desired direction. One group (7%) used the index thumb and middle finger, which by joining together can proceed to move the page in the required direction. The most popular gesture involves the use of an open hand; however, this gesture is not considered for this action due to the current configuration of the dictionary. Until now, all actions have been performed using the index finger, and we aim to maintain consistency for fundamental gestures to be as agile and intuitive as possible. While the user is already moving the cursor through the index, it is desirable to add the middle finger to activate the scrolling functionality, configuring it as the definitive gesture, see Fig. 1. Page resizing. The page resizing gesture for the page comprised two distinct actions: zoom in and zoom out. In each designer group, the creation of the zoom-in gesture was consistently followed by the development of a contrasting gesture to facilitate zoom-out, ultimately restoring the page to its default size. Out of the 16 groups studied, five groups (31%) opted for a gesture involving both hands initially placed adjacent to each other, followed by expanding them outward. Another group (6%) proposed a similar version, beginning with hands touching together. Four groups (25%) selected the opening of the index finger and thumb. Another 4 groups (25%) chose the simultaneous opening of all fingers to execute the zoom-in action. One group (6%) utilized the opening of the index, middle finger, and thumb for zooming in. A different group (6%) preferred starting with all fingers closed, and gradually opening the hand while moving closer to the display. Lastly, one group (6%) employed both hands, joining the thumb, middle, and index fingers, with their expansion triggering the zoom-in action.

Figure 2: Page resizing; Browsing history; Panning gestures.

The most widely adopted gesture for resizing the page involves the expansion of both hands, as illustrated in Fig. 2. Recognizing that the action of resizing the page is not as frequent as pointing or confirmation, the justification for employing a different hand/finger configuration for this specific task is justified.

Quick browsing history. Three groups (19%) did not propose this gesture. From the remaining 13 groups, the majority, 9 groups (69%), opted to navigate history by employing the entire hand to make a quick movement to the left. One group (8%) chose a rapid movement to the left using their hand, specifically with the thumb raised and the other fingers closed. Another group (8%) decided on a quick movement by shifting the right elbow to the left. Yet another group (8%) envisioned a clockwise circular movement of the entire right hand. Finally, one group (8%) used two fingers to swipe to the left. Among those who created this gesture, 5 groups (38%) created a gesture exclusively to go back to the recently visited page. Meanwhile, 8 groups (61%) created a gesture for both backward and forward navigation in history. For these groups, the gesture for moving forward is identical to the one for moving backwards, but the direction of the movement is reversed. The most frequently recurring gesture for navigating in history involves using the entire hand to make a swift movement to the left or right, depending on the desired direction in the browsing history. This gesture (see Fig. 2), being the most popular (69%) within the sample, has been chosen for inclusion in the final dictionary. The movement for navigating in history should be completed within a timeframe of $t <= 2s$. This timing is justified by the shared use of the open hand gesture for various other actions, as will be discussed later. Establishing these time limits is essential to ensure accurate identification of each action by the system.

Panning in resized page. Only 5 groups (31%) out of 16 choose to include the action of panning in a resized page when the zoom of the viewport is >100%. 3 groups (60%) use the following gesture sequence: beginning with an open hand closing where no buttons or links are present, then moving the hand will correspond to moving the webpage, followed by releasing the hand. 1 group (20%) uses the middle, index, and thumb fingers of the right hand to drag the page in the desired direction. 1 group (20%) decided to use the index and middle finger where their movement will correspond to dragging the page. The gesture selected for panning within a resized page involves the closure of the hand in the whitespace, see Fig. 2.

Drag and drop. Despite being implemented by only 5 groups; all of them (100%) chose a common gesture combination: closing the hand over the desired element and then moving the hand to the destination before releasing it. Although this gesture is the same as the one used for panning within a page, the system can distinguish between them. The drag-and-drop gesture requires that the hand is closed on a webpage element. In contrast, for the panning action, the hand must not be closed on buttons or links but only in the whitespace. This specification validates the inclusion of this combination of gestures in the final dictionary.

Interaction initialization and stop. Even though only 2 out of the 16 groups (13%) implemented the gesture to initiate the interaction, this task is considered crucial, particularly in a public context. Therefore, it will be included in the final dictionary. One group out of the two (50%) replicated the "greeting gesture" by moving the hand quickly from left to right. The other group (50%) proposed using the open hand, waiting for a duration of $t >= 3s$ in front of the display to initialize the system. Both approaches contribute to the initiation of the system and provide users with options for starting the touchless interaction. One of the groups that implemented the gesture for initiating the interaction also introduced a corresponding gesture for ending the interaction. This gesture involves the use of an open handheld for a duration of $t >= 4s$, as shown in Fig. 3.

Figure 3: Initialization and Stop; Reload; Homepage access (1-2).

To start the interaction, in the final dictionary we incorporated the "greeting gesture", namely moving the hand quickly from left to right and vice versa. This is a spontaneous and shared gesture, also proposed in other touchless interactions [13]. To stop the interaction, users must hold a still hand for $t >= 4s$. This design allows the gesture to work seamlessly at any phase of web navigation, providing users with the flexibility to exit the interaction at their discretion.

Page reload. Only one group (6%) out of the 16 introduced a gesture to refresh the page. The proposed gesture involves pointing the index finger towards the display and making a complete clockwise rotation of the index finger, forming a spiral, as shown in Fig. 3. The completion of the spiral triggers the web page to reload.

Homepage access. Only 1 group out of 16 (6%) created a gesture to quickly return to the homepage. To execute it, both hands need to be raised and rotated by 90° so that both palms face each other, then quickly join them together. The gesture will take the user to the homepage. This gesture is naturally distinguishable from others, thus avoiding potential conflicts and being eligible for the final dictionary.

3 CONCLUSION AND FUTURE WORK

In this paper, we presented the results of an empirical evaluation of mid-air gestures, carried out with 56 students who acted as designers and 273 external participants, who were involved either as co-designers (25) or as evaluators of the proposed gestures (248). Our main contribution is the definition of a dictionary of mid-air gestures optimized for web browsing and tested during an interaction with a simulated university classroom search service. In addition, our study allowed us to observe recurring behaviours which might be interesting also for designers of touchless interfaces who are approaching different domains. We found that users tend to adopt gestures primarily designed for touch-based and mouse-based interfaces, such as those we can find in most mobile devices, and also for touchless interactions. The problem with this approach is that these gestures can be inefficient when they are used for more complex tasks in a touchless environment, leading to misinterpretation of user intent. One of the reasons which might have led the participants in our study to devise and/or positively assess such gestures is that laptops were used to simulate touchless devices. Consequently, all gestures were performed at a relatively small distance (approximately 50 cm) from the screen, an unlikely condition when interacting with large displays.

In future work, we deem it important to further evaluate the dictionary we have proposed. Firstly, we aim to test the proposed gestures in a more realistic context, thus actually using large displays. Secondly, we need to verify that the adjustments and additions we have made to the original gestures identified by the designers (e.g., regarding timings) do not compromise their guessability.

A limitation of our study is that designers focused on a single specific context, i.e., interaction with a university classroom search service. Thus, the dictionary does not include actions which were not considered relevant for this scenario, either by the original designers or by the authors. Hence, it might not be possible to transpose the dictionary to a different web context without any additions and modifications.

Finally, we realize that, if large touchless displays were to be installed on the university campus, it would be advisable to make a tutorial for first-time users easily and always accessible, thus allowing them to practice mid-air gestures and gain familiarity. In fact, building confidence in the end user is crucial to minimizing potential errors in gesture enactment.

REFERENCES

[1] Salvatore Andolina, Paolo Ariano, Davide Brunetti, Nicolo Celadon, Guido Coppo, Alain Favetto, Cristina Gena, Sebastiano Giordano, and Fabiana Vernero. 2019. Experimenting with Large Displays and Gestural Interaction in the Smart Factory. In *2019 IEEE International Conference on Systems, Man and Cybernetics, SMC 2019,*

Bari, Italy, October 6-9, 2019. IEEE, 2864–2869. https://doi.org/10.1109/SMC.2019.8913900

[2] Carmelo Ardito, Paolo Buono, Maria Francesca Costabile, and Giuseppe Desolda. 2015. Interaction with Large Displays: A Survey. *ACM Comput. Surv.* 47, 3 (2015), 46:1–46:38. https://doi.org/10.1145/2682623

[3] Aiswarya Babu, Zahiriddin Rustamov, and Sherzod Turaev. 2023. INTELLIGENT TOUCHLESS SYSTEM BASED ON GESTURE RECOGNITION. *Journal of Theoretical and Applied Information Technology* 101, 10 (2023), 3936–3942.

[4] Kathy Baxter, Catherine Courage, and Kelly Caine. 2015. *Understanding your users: a practical guide to user research methods*. Morgan Kaufmann.

[5] Bibiana Bayer, Kerstin Blumenstein, Thomas Ederer, Stefanie Größbacher, David Mayerhuber, Sabrina Rockenschaub, Grischa Schmiedl, and Carina Skladal. [n. d.]. Shop Window Control. ([n. d.]).

[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[7] Davide Brunetti, Cristina Gena, and Fabiana Vernero. 2022. Smart interactive technologies in the human-centric factory 5.0: a survey. *Applied Sciences* 12, 16 (2022), 7965.

[8] Sugnik Roy Chowdhury, Sumit Pathak, and MD Anto Praveena. 2020. Gesture recognition based virtual mouse and keyboard. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. IEEE, 585–589.

[9] Luca Console, Fabrizio Antonelli, Giulia Biamino, Francesca Carmagnola, Federica Cena, Elisa Chiabrando, Vincenzo Cuciti, Matteo Demichelis, Franco Fassio, Fabrizio Franceschi, Roberto Furnari, Cristina Gena, Marina Geymonat, Piercarlo Grimaldi, Pierluigi Grillo, Silvia Likavec, Ilaria Lombardi, Dario Mana, Alessandro Marcengo, Michele Mioli, Mario Mirabelli, Monica Perrero, Claudia Picardi, Federica Protti, Amon Rapp, Rossana Simeoni, Daniele Theseider Dupré, Ilaria Torre, Andrea Toso, Fabio Torta, and Fabiana Vernero. 2013. Interacting with social networks of intelligent things and people in the world of gastronomy. *ACM Trans. Interact. Intell. Syst.* 3, 1 (2013), 4:1–4:38. https://doi.org/10.1145/2448116.2448120

[10] Roshnee Matlani, Roshan Dadlani, Sharv Dumbre, Shruti Mishra, and Abha Tewari. 2021. Virtual mouse using hand gestures. In *2021 international conference on technological advancements and innovations (ICTAI)*. IEEE, 340–345.

[11] Pinku Deb Nath, William Delamare, and Khalad Hasan. 2024. PalmSpace: Leveraging the palm for touchless interaction on public touch screen devices. *International Journal of Human-Computer Studies* 184 (2024), 103219.

[12] Atul Kumar Prasad, Akshat Sharma, et al. 2022. Gesticulation Recognition System Using Deep Learning. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, 1000–1002.

[13] Kabid Hassan Shibly, Samrat Kumar Dey, Md Aminul Islam, and Shahriar Iftekhar Showrav. 2019. Design and development of hand gesture based virtual mouse. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*. IEEE, 1–5.