User-Based Usability Testing Report

Names of All Group Members

- Seçil Gezer 29145
- Sevdenaz Yılmaz 30300
- Deren Doğan 29561
- Elifnur Zengin 31237

1. Methodology

Users:

Six users participated in the usability test. The participants ranged in age from 22 to 24, with three females and two males. None of the participants had prior experience with gesture-based applications. Diverse academic and professional backgrounds were ensured.

User ID	Age	Gender	Gesture-Based Experience	Gaming Experience	Department
U1	24	Male	None	High	CS
U2	23	Female	None	None	CS
U3	22	Female	None	High	CS
U4	22	Female	None	None	IE
U5	22	Male	None	None	EE
U6	22	Female	None	None	CS

Test Procedure:

- **Before the Test:** We explained the purpose of the test to the participants. We provided information about the tasks they would perform while interacting with the application and what these tasks were testing. We explained to the participants that their purpose was only to evaluate the functionality of the application and that any mistakes they made would not harm the testing process. We explained to the participants what the "think aloud" protocol was and how to apply it. We asked the participants to express their thoughts out loud while interacting with the application. We explained that this protocol would help us understand their decision-making processes and the challenges they faced. We told the participants that they should express their thoughts at each step by asking themselves questions such as "What is this button?" and "Why am I choosing this?" We explained that the data collected during the

test would only be used to evaluate the test and that confidentiality would be ensured. We informed the participants that their personal information would be recorded anonymously and used only for research purposes. We stated that it was not mandatory for their faces to be visible in the photo shoot during the test. Participants were informed about the stages of the test, which would first be a brief introduction about the application, then the tasks to be performed during the user test, and finally a post-test interview and SUS survey.

Figure 1: *User Testings*

- **During the Test:**
  Participants were asked to interact with the application independently to complete specific tasks. The tasks included menu navigation, item selection, and basic interactions with the user interface. Each participant was given tasks such as selecting an item in the menu or performing a specific action. Participants shared their thoughts by verbalizing each step according to the “think aloud” protocol. This helped us understand what choices they made and what steps they followed during the task. During the test, users’ interactions were carefully observed and every action they took was recorded. In addition, users’ facial expressions, body language, and reactions were also observed to collect additional data about the experience. Participants who encountered any difficulties were asked how they tried to overcome these obstacles. The interaction time was also recorded to evaluate the overall effectiveness of the test. Careful notes were also taken on whether the application correctly detected the items in the menu and whether the transitions were smooth. It collected important data on how the user interacted with the interface and what difficulties they encountered.

- **After the Test:** Every participant took part in a post-test interview following the test and shared their feedback on the experience. In the interview, participants were prompted to elaborate on their experiences with the app, which features they found beneficial or difficult, as well as any issues they faced with the user interface, and observations were recorded. Participants were inquired about aspects like which features felt the most intuitive, which were puzzling, and the overall user-friendliness of the app. This was a crucial move in collecting users' subjective feedback and gaining a more profound insight into the app's user experience. Participants additionally filled out the System Usability Scale (SUS) survey. This questionnaire is a typical approach for assessing user experience and intends to gauge the user's overall contentment. At the conclusion of the survey, every participant was requested to evaluate the app's overall usability. The SUS assists users in assessing the efficiency, effectiveness, and satisfaction of the interface. The survey results underwent statistical analysis to assist in pinpointing the app's strengths and weaknesses.

Task Procedure:

Participants were guided to complete specific tasks while interacting with the application. Each participant was informed at the beginning of the test that they would need to express their thoughts out loud according to the "think aloud" protocol and were asked to perform the tasks independently. At the beginning of the test, participants were given a briefing explaining the five basic gestures they would be using (Point Up, Point Down, Swipe Left, Swipe Right, Thumbs Up) and what functions these gestures corresponded to. Participants began the initial phase of the test by standing before the device and arranging their hands correctly. At this stage, they were requested to carry out a calibration movement to enable the system to accurately recognize their hand motions.

Menu Navigation

The initial task involved moving through the menu in a vertical direction. Participants were directed to choose items in the menu by using the "Point Up" and "Point Down" gestures. The system indicated the chosen item by moving it up or down using these gestures. Participants noticed how precisely the gestures were recognized while they moved through the options in the menu.

Navigating Between Columns

In the second task, participants were asked to navigate between columns in the menu. They were asked to move the cursor to the right or left using the "Swipe Left" and "Swipe Right" gestures. During this task, participants attempted to navigate between different columns in the menu using both gestures.

Item Selection

In the final task, participants used the "Thumbs Up" gesture to select the selected item. When the system detected this gesture, it highlighted the selected item in green and displayed the message "Clicked!". Participants repeated this selection process several times to test the accuracy of the gestures.

Context: The research showcases an application trial that allows users to engage with gestures, such as navigating menus and performing selection tasks. The participants are people experiencing this system for the first time and lack prior experience with similar gesture-based interactions. The test is carried out to assess the system's intuitiveness and ease of use. It also examines how users view this gesture-based system overall and what challenges they face during interaction. The aim of the test is to gain insights into the overall usability of this system and its interaction processes while gathering feedback for future improvements.

Tool:

The tests were conducted using a **Macbook Air** device with the following specifications:

- Processor: Intel Core i5
- RAM: 8GB
- Operating System: macOS
- Application: Gesture-based interactive system.

2. Results

Qualitative and Quantitative Data:

User ID	Positives	Negatives	Common Themes	Critical Themes	Solutions	Quantitative Data
U1	Gestures were intuitive and responsive.	Item selection was difficult, and it was hard to navigate with arrow movements.	Intuitive gestures but challenges with item selection.	Difficulty in selecting items accurately.	Improve gesture recognition for item selection.	Test Duration: 4 min
U2	Easy to understand vertical and horizontal navigation.	Item selection was difficult.	Smooth vertical navigation but issues with item selection.	Issues with precise navigation using arrow gestures.	Improve the accuracy of arrow-based navigation.	Test Duration: 5 min

U3	Horizontal navigation felt smooth.	The system sometimes confused Thumbs Up with Point Up.	Clear navigation feedback but gesture misinterpretation.	Misinterpretation of Thumbs Up as a Point Up gesture.	Improve gesture detection to avoid misinterpretation.	Test Duration: 4 min
U4	Thumbs Up selection provided clear feedback	Item selection was difficult, and it was hard to navigate with arrow movements.	Item selection process was challenging.	Difficulties with precise item selection.	Optimize item selection process and improve gesture recognition.	Test Duration: 5 min
U5	The gesture-based system felt innovative.	The system sometimes confused Thumbs Up with Point Up.	Innovative but misinterpretation of key gestures.	Misinterpretation of Thumbs Up as a Point Up gesture.	Improve the detection algorithm to distinguish between Thumbs Up and Point Up.	Test Duration: 4 min
U6	Simple gestures made learning easy.	Item selection was difficult, and it was hard to navigate with arrow movements.	Gestures were simple but navigation with arrows was difficult.	Challenges in item selection and navigation.	Provide a more intuitive navigation method for item selection.	Test Duration: 4 min

Quantitative Data:

- **Number of Users**: 6
- **Average Time Spent**: Approximately 5 minutes per participant.
- **Common Errors**:

  1.The Thumbs Up gesture was sometimes misrecognized as Point Up.

  2.Difficulty navigating the menu using arrow gestures.

  3.Delay in the system's response during item selection.

- **SUS Scores:** The average SUS score was calculated as 70, indicating that the system is moderately usable.

User ID	SUS Score
U1	70
U2	67
U3	75
U4	65
U5	75
U6	68

Post-Test Interviews:

1. What did you like about the application?

U1: The gestures were intuitive and responsive.

U2: I found the vertical and horizontal navigation easy to understand.

U3: I liked the smooth horizontal navigation.

U4: The visual feedback for the Thumbs Up gesture was clear.

U5: The gesture-based system felt innovative.

U6: The gestures were simple and easy to learn.

2. What difficulties did you experience while using the application?

U1: I had trouble selecting items; the gesture recognition wasn't always accurate.

U2: The arrow gestures (Swipe Left and Swipe Right) were inconsistent and hard to control.

U3: The system sometimes misrecognized the Thumbs Up gesture as Point Up.

U4: The pointer moved too quickly, making precise selection difficult.

U5: The system had slight delays when recognizing gestures.

U6: It was hard to maintain a stable hand position for accurate gesture recognition.

3. How can the system be improved?

U1: Improve the gesture recognition speed and accuracy, especially for item selection.

U2: Adjust the sensitivity for navigation gestures like Swipe Left and Swipe Right.

U3: Enhance the accuracy of Thumbs Up detection to avoid misinterpretation.

U4: Allow customization for pointer speed to improve precise selection.

U5: Optimize the system response time to reduce delays.

U6: Provide visual guides or indicators to help maintain hand stability during gestures.

Post-test interviews reveal that while participants appreciated the intuitive and aesthetic design of the gesture-based system, they still encountered usability challenges related to gesture recognition and navigation. Misinterpretation of gestures, such as Thumbs Up being interpreted as Point Up, significantly impacted the user experience. Additionally, issues with gesture accuracy and system responsiveness highlighted areas for improvement. Participants' feedback highlighted the need for improved gesture accuracy, better system responsiveness, and improved navigation control, providing valuable insights into improving the application to deliver a smoother and more user-friendly experience.

3. Discussion and Conclusion

1)What Data Tell Us About the Application

The results of the usability test indicate that the gesture-based interactive system is moderately usable, with a **System Usability Scale (SUS) score of 70**. Participants found the gesture-based interactions innovative, intuitive, and easy to learn particularly for vertical and horizontal navigation tasks. Positive feedback during the post-test interviews highlights that users appreciated the system's design and visual feedback, especially for gestures like Thumbs Up. However, usability issues remain particularly in tasks involving item selection and navigation accuracy.

The quantitative data further supports these findings:

- Common errors such as misinterpretation of the Thumbs Up gesture as Point Up, were observed.
- Average task completion times (4–5 minutes) reflect that while users could perform tasks, they faced challenges with gesture accuracy and smooth transitions.Additionally, navigation sensitivity and gesture delays made precise control difficult, particularly when using Swipe gestures for horizontal transitions.

Usability Issues:

2)What Are the Usability Issues? Why?

- **Gesture Recognition**

  Complex gestures, such as Thumbs Up, were not consistently recognized and were sometimes misinterpreted as Point Up because the gesture detection algorithm lacked precision and failed to distinguish between similar gestures effectively. This inconsistency reduced user performance and increased frustration, particularly during tasks requiring accurate item selection.

- **Feedback Mechanisms**

  Lack of clear visual feedback for gestures like Swipe Left/Right and Point Up/Down made it difficult for users to understand whether their gestures were successfully recognized.The system did not provide sufficient visual indicators, such as highlights or confirmation messages, for these gestures. This absence caused confusion and led users to repeat gestures unnecessarily.

- **Learning Curve**

  Some users struggled to learn and remember the basic gestures, particularly during the initial phase of testing, especially those without prior experience with gesture-based systems.The absence of clear onboarding instructions or real-time visual aids contributed to the difficulty in learning the required gestures

- **Item Selection Difficulty**

  Selecting items using gestures was requiring multiple attempts to complete the task because delays in system response and inaccurate gesture detection for selection gestures, such as Thumbs Up, made the process challenging. This resulted in increased task completion time and frustration for users. Some users struggled to learn and remember the basic gestures, particularly during the initial phase of testing.

- **Navigation Challenges**

  Users experienced difficulty navigating the menu with Swipe Left/Right gestures due to the system's inconsistent and sensitive behavior. The lack of control and stability in the pointer's movement resulted in imprecise navigation.Users were unable to smoothly transition between menu columns, which hampered task performance.

- **Hand Stability Issues**

  Users found it hard to maintain a stable hand position during gesture input.The system's sensitivity to small hand movements caused errors in gesture detection. A lack of visual guides or feedback worsened this issue, as users were unsure how to adjust their hand positioning.

3) How can you improve the system based on the feedback from the users?

Some improvements could enhance the system's usability and user experience. One of the problems under discussion is gesture recognition with notable misinterpretation cases within THUMBS UP wherein the gesture has been interpreted as POINT UP resulting in errors in choosing the items. For tackling the above issue, one can adopt a method based on machine learning, like using a gesture recognition model of the convolutional neural network type, and initiating the use by engaging a gesture calibration stage at the start which enables the setting to learn each user's interpretation of hand movements. In addition to this, use confidence scores provided by MediaPipe to trigger gestures only when they exceed the confidence threshold of 90% and allow change of gesture sensitivity which would improve accuracy.

One other important point was the user oblivious condition where real-time feedback was left. This could instead show to the user how well the gestures were being registered. The above may be done with real-time gesture show indicators. These indicators should be tied at the top of the screen in conjunction with some visual cues (such as highlighted or colored items indicating selection) that may prompt the user that their action has been successful. An audio cue might also be part of feedback, such as the usual sound made by clicking when one chooses an item. Swipe Left/Right gestures can use swiped trail animation to smoothen their user experience; that way, it becomes intuitive to the users.

Users have also said that they experienced a very steep learning curve, especially for new users who had never been familiar with gesture-based systems. This improvement could be made with an interactive onboarding tutorial, including animated gesture demonstrations. Idle-time hints and a help button offering gesture guidance would further support users. One possibility is a "hold-to-confirm" mechanism that will let the users hold the Thumbs Up gesture for one second before confirming that they mean it with a progress bar showing how much time is left to make the selection; this would mean less effort is needed for item selection. Optimizing system requirements so that gestures could be recognized swiftly would bring the delays in selection down.

The last concern was about various hand stability issues since the small movement of hands sometimes led to a reconsideration of concepts. A hand guidance system could tell users where to keep their hands while a hand alignment algorithm could automatically cater to small deviations. Less sensitive gesture features suited to accommodate hand movements that are less stable would also be helpful. Each of these improvements would significantly enhance the usability of the system, reduce frustration, and make the overall experience feel seamless and friendly.

4) Findings related to HCI literature

Usability tests revealed findings in line with some of the principles of Norman's Seven Stages of Action and Shneiderman's Eight Golden Rules of Interface Design, which emphasize clear feedback and intuitive design and learning ease. The major problem seen was the ambiguity in gesture interpretation, especially between the gesture of "Thumbs Up" and "Point Up". This has direct reference to Norman's (Norman, 2013) "Gulf of Execution", where the user knows the action they intend to perform (selecting an item) but has trouble executing it. They unsuccessfully attempt it because the system will not recognize the intended gesture. Better detection accuracy and real-time feedback reduce this gulf, thus making the interaction smoother.

Another major problem was that gestures were indicated as accurately recognized; however, the user would repeat them because there was no feedback. As per Shneiderman's Golden Rule of Feedback (Shneiderman et al., 2016), the system must supply non-stop and immediate feedback to user actions. However, in the system, there are no visual or audio confirmations for successful gestures, showing that this principle has been violated. Users will gain confidence that their acts have been registered, thus minimizing errors and increasing task efficiency by incorporating visual cues, such as item highlights, and audio clues, such as sounds of clicks (Norman, 2013).

Findings have also indicated that most users have difficulty in learning and remembering gestures, especially during the onset of the interaction. This is in line with Norman's concept of "knowledge in the head" versus "knowledge in the world" (Norman, 2013). According to this argument, users are not supposed to carry the memory of using specific system instructions. To achieve this, using interactive onboarding tutorials and gesture hints on the screen will act as "knowledge in the world," where gesture instructions are made accessible at the point of need. This also supports Shneiderman's rule of "help and documentation", as users can access guidance whenever required (Shneiderman et al., 2016).

Another important aspect is the hand stability problem, characterized by very small movements of the hand that lead to misinterpretation by the system of gestures. This can be linked to Norman's principle of Affordance, which says that the users should understand naturally how to interact with an object. In this case, lack of proper affordance for hand placement fails to show users how they should position their hands. As a solution, the system for guidance of hand position can provide the visual indicator, where the users would have to place their hands to improve affordance and decrease errors. Moreover, giving power to the user to change his or her parameters on gestures would comply with Shneiderman's flexibility and user-control principles so that the system adapts towards the user's physical condition.

It is finally established that these delays are associated with Norman's Gulf of Evaluation, where users have difficulty at times determining whether an action has been effective. When input gestures are not rapidly processed, an end-user could easily get displeased by feeling as though a behavior was not being captured by the system. Shneiderman's principle of keeping such response time as minimum conceivable emphasizes never subjecting the user to longitudinal waits to prevent interruption of the flow. This means somehow modifying the glitch detection even further for quick input expressiveness, besides adding the common guidance such as progress bars in gesture recognition, really brightens understanding without ambiguity.

Briefly, they reveal the crucial absence of feedback, learning, control, and system response time, all of which are addressed directly in Norman's Seven Stages of Action along with Shneiderman's Eight Golden Rules. The system design will be enriched eventually through enjoining

such terms like feedback, affordance, and learnability so that a more intuitive, efficient, and friendly user experience can be provided.

5) Limitations of the study

The usability testing had to acknowledge some limitations. Firstly, the restricted size of the sample as only six users were involved. This stops the generalization of the findings since the sample would not necessarily represent users with different motor skills, cognitive abilities, and experience with technology (Nielsen, 1994). Most of the participants are, however, university students with a technical background, thus creating a demographic bias. Users from non-technical fields or of older age groups might interact with the system differently and suggest that future studies should have a more diverse participant pool (Lewis, 2014).

Another limitation of the study is the testing conditions: all experiments used the same environment - quiet, distraction-free, and same device- a MacBook Air. Consistency was, however, ensured at a cost of not accurately simulating real-world conditions, such as noise distractions or varying lighting (Norman, 2013). Indeed, only such a setting could provide more realistic insights. Furthermore, testing across devices creates the problem of device dependency, as gesture recognition may differ because of variations in camera resolution, screen sizes, and hardware specifications - this is the case with MacBook Air and others as well (Shneiderman et al., 2016). Testing smartphones, tablets, and other devices would provide a more holistic view of system performance.

Ultimately, there is no precise error analysis, as although errors of interpretation of gestures (e.g., Thumbs Up versus Point Up) have been noted, actual error rates have not been calculated. An accurate figure measuring the percentage of errors per gesture would define more working improvement aspects to the system (Lewis, 2014). The research is considered because other environmental factors such as lighting changes, camera quality, or other moving foreground effects that would also make a difference to recognition are not considered in this research (Norman, 2013). Testing into various conditions would thus assure the reliability of this system in real-life environments. Addressing such would present a more detailed scope of the user's performance in using this system into a more flexible, user-friendly product.

4. References

Lewis, J. R. (2014). *Usability: Lessons Learned… and Yet to Be Learned*. International Journal of Human-Computer Interaction, 30(9), 663–684.

Nielsen, J. (1994). *Usability Engineering*. Academic Press.

Norman, D. A. (2013). *The Design of Everyday Things* (Revised and expanded edition). Basic Books.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., & Elmqvist, N. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th ed.). Pearson.