Yatri Guide: A Touch-Free Interface for Public Transport Assistance

Suyash Chintawar - 191IT109
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
Email: suyash.191it109@nitk.edu.in

Naveen Shenoy - 191IT134
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
Email: naveenshenoy.191it134@nitk.edu.in

Navaneeth P - 191IT132
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, India 575025
Email: navaneethp.191it132@nitk.edu.in

***Abstract*—A very large proportion of the interfaces used in day to day life involve the use of a cursor to navigate over the screen. These interfaces necessitate the use of the hand to physically move the mouse thus moving the cursor on the screen. This requirement on the user side may not always be suitable or convenient. In an environment with restriction on motor movements, the option of having facial movements to control the interfaces becomes very useful. We propose a public transport assistance application driven with cursor control using facial gestures. The application supports checking availability of express trains, buses and local trains along with their fares between stations. The application is targeted towards users who are not in a situation to use their hands to use a cursor. This includes motor disabled people or people who may not want to touch a public interface due to hygiene concerns.**

Key words— HCI, transport assistance guide, motor disabled, cursor control, facial gestures

I. Introduction

As we move towards a more technologically advanced world, there is a rapid increase in the use of intelligent public interfaces to guide humans in their day to day activities. Such interfaces have appeared in many domains such as public transport, entertainment, and e-commerce. Such applications in the public domain should have enhanced accessibility so that no section of the population misses out on the benefits of the new and advanced technology. A variety of possible use cases of the application should be considered in the design phase to increase its usability. In the design process of such an application, it is necessary to keep the user involved in the form of taking continuous feedback. HCI principles need to be applied at each step keeping the developments in the interests of both novice and experienced users.

Human gesture recognition systems have gained popularity as of late due to the advancements in the field of deep learning. Robust deep learning networks can be used to effectively recognise human gestures and perform custom actions forming the basis of human computer interaction. The developments in the public transport sector can be further complemented by the inclusion of smart public transport assistance interfaces which can guide users with their travel plans. The key contributions of our work is a novel public transport assistance system based on cursor control using facial gestures. In the public transport landscape, a passenger may want access to data to guide their travel, although their hands may be preoccupied due to carrying luggage or due to any motor disability. This is where our application, Yatri Guide, comes in. It allows touch-free access to a transport assistance interface which can guide users regarding the availability and fares of upcoming trains and buses. The following sections of the paper are as follows:- In section II, we discuss about the various previous work done in this domain. After expressing the problem statement along with the objectives in section III, we move onto the methodology in Section IV. We describe the HCI principles applied in our work in section V and in section VI, we show the experimental results and analysis of our work. Finally, we discuss an overview and future scope of this research in section VII.

II. Literature Survey

The authors of this paper [1] proposed and built a hands free device which controls the computer cursor functionalities using a person's head movements and their cheek muscle twitches. Results obtained here showed that it took 2.37 seconds on average to move the cursor from one point to the other on the screen. Almost double the time was taken compared to the the use of an traditional optical mouse to perform the same task. 5.28% false negative clicks were recorded. EMG sensors were used by the authors of this paper [2] to control the cursor movements. Two different applications, one with auto rotate functionality and other with manual rotation were built and the user experience was compared and analysed. 14 different users were made to use both of these applications. A Fitt's law task was given to the users for evaluating the

TABLE I
SUMMARY OF LITERATURE SURVEY

Authors	Methodology	Merits	Limitations
K. Sancheti et al. [1]	Cursor Control using head movements and cheek muscle twitches	High accuracy and low false negative clicks	Almost double time consumed for performing actions
O'Meara et al. [2]	Use of EMG Sensors for cursor control	Detailed user experience analysis	Only one Fitt's law task used
P. Kopp. [3]	Analysis and improvement of facial landmark detection	3D landmarks stay in correspondence with their physical positions on the face	Training data was limited and frames in a sequence were highly correlated
H.S. Grif [4]	Proposed a novel hand gesture based cursor control system	Reliable and accurate	Evaluation metrics could be better
S. Mathew [5]	Cursor control using eye movements	Real time eye-tracking	Actual user experience analysis missing

systems. The results showed that over time, the throughput and path efficiency increased and the no. of overshoots decreased. The study [3] talks about the recent work in detecting facial landmarks. It focused on evaluating the entire process of landmark detection including face detection. It was found that box-initializations play an important role in detecting facial landmarks. Regression tree methods perform outstandingly on low resolution sequences. On the other hand, CNN based facial landmark detectors performed better on high resolution inputs. A hand gesture based cursor control system was proposed in [4]. This paper used the human hand as a mouse by mapping hand gestures to the optical mouse functionalities. The system was tested under different weather conditions to study the difference in levels of illuminance. The authors of the paper [5] have proposed an eye-movement based cursor control for the specially abled citizens. dlib's 68 points shae predictor was used. The main drawback of this paper included the lack of empirical analysis and missing actual user experience.

III. Problem Statement

To develop a cursor control system controlled by facial movements for enhancing user accessibility.

A. Objectives

- To create a cursor control system based on facial movements.
- To choose and map facial gestures to actions and test their different activation thresholds.
- To build a Public Transport Assistance system to showcase the cursor control system.
- To analyse the gestures and different aspects of the system using HCI principles and guidelines.
- To obtain feedback on various metrics to analyse the overall system.

IV. Methodology

The flow of the proposed cursor control system can be seen in Fig. 1. The built cursor control system supports activation as well as deactivation using only one facial action, that is, opening of the mouth. The major steps involved in the flowchart are explained below,

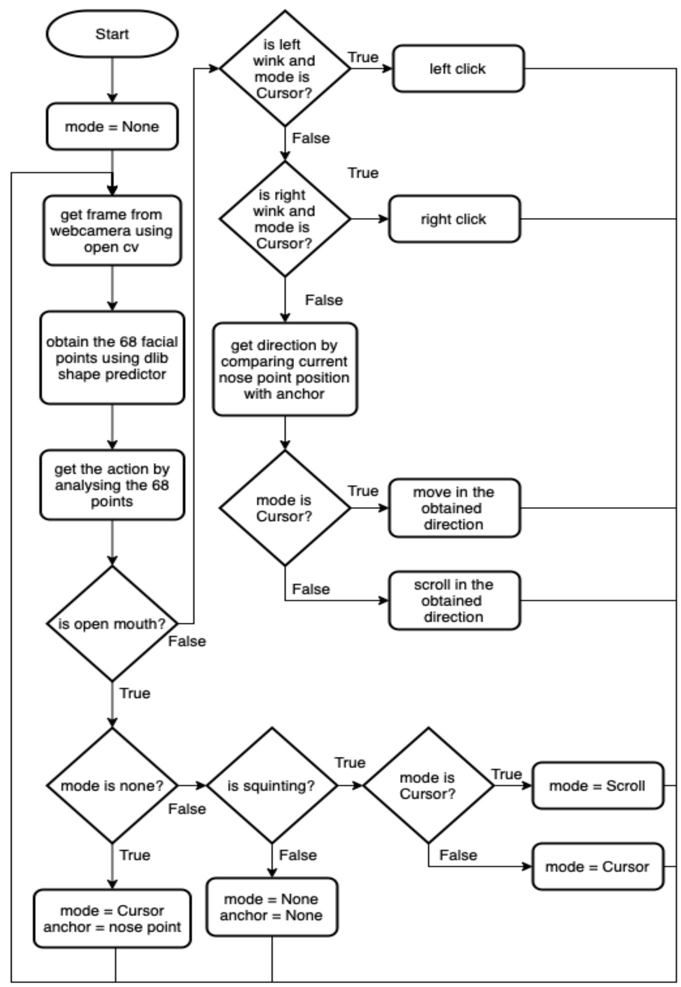

Fig. 1. Flowchart of the cursor control system

A. Facial recognition

The first step involves recognising the position of different features of the face like eyes, nose, mouth and eye brows. Dlib's `shape_predictor_68_face_landmarks` [6] uses an ensemble of regression trees approach to provide 68 points covering all these facial features. Fig. 2 shows the facial landmarks extracted from dlib. The following table lists the facial feature and the indices within the 68 points,

Facial feature	Indices
Jaw	[00, 17)
Right eyebrow	[17, 22)
Left eyebrow	[22, 27)
Nose	[27, 36)
Right eye	[36, 42)
Left eye	[42, 48)
Right eye	[48, 68)

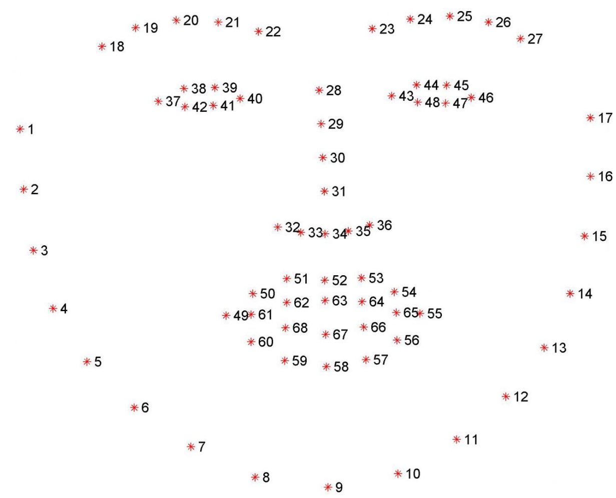

Fig. 2. 68 points of the face obtained by dlib shape predictor

B. Facial Action recognition

Once the positions of the facial features are extracted, the next stage constitutes of taking the 68 facial points as input and emitting of an action. Initially the following four actions were chosen,

- Mouth open
- Left wink
- Right wink
- Squinting

After several experiments, it was discovered that looking at the bottom of the screen would trigger a squinting action contributing to a false positive action. The squinting action had to be made more intentional since looking at the bottom of the screen was a normal eye movement. Eventually mouth open was chosen as an activation gesture along with squinting. Therefore the four actions employed in the system are,

- Mouth open
- Left wink
- Right wink
- Squinting with Mouth open

An obstacle in the smooth detection of winking recognition was the fact that the model was never trained on the faces with winks which lead to the difference in eye aspect ratio being very small. Having employed this small difference as an activation for winking, it was found with repeated experiments that turning the head left or right would also activate winking. Looking closer to the the eye brow movements while winking, it was observed that there was a significant difference in their distance from their respective eye. Using this trait as another activation step for winking resolved the conflict.

C. Accessibility

With the action emission complete, the next step was to map the facial action to a cursor action. The list of cursor actions that were targeted are as follows:

- Toggle cursor system activation (1 action)
- Left and right click (2 actions)
- Move left, right, up and down (4 actions)
- Scroll left, right, up and down (4 actions)

The following mapping was chosen to convert the facial action into a cursor action:

Facial action	Cursor action
Mouth open	Toggle cursor system
Left wink	Left click
Right wink	Right click
Squinting with mouth open	Toggle mode

Different modes have been introduced to switch between scrolling and cursor movement:

- Cursor Mode
- Scroll Mode

When the cursor system is activated by opening the mouth, the position of the tip of the nose is saved as anchor. The direction of scroll or cursor movement is determined by the current position of the tip of the nose with respect to the anchored point. Therefore when the direction is right, the cursor moves right in cursor mode and scroll right in scroll mode. The mode can be toggled by squinting with mouth open action.

The mapping of head movement to cursor movement in Cursor Mode is shown below:

Head movement	Cursor action
Look left	Move left
Look right	Move right
Look up	Move up
Look down	Move down

The mapping of head movement to scroll direction in Scroll Mode is shown below:

Head movement	Scroll direction
Look left	Scroll left
Look right	Scroll right
Look up	Scroll up
Look down	Scroll down

There is a possibility that the user might hold the gesture for a longer period of time, thereby triggering the same action several times. This was resolved by introducing a NONE action and deactivating an action for a few frames after it's been performed.

Each gesture to be performed by the user to achieve a certain goal has been mapped with utmost consideration. The cursor system is activated and deactivated by opening the mouth. The reason for choosing this action was that opening of mouth is not an action that is performed subconsciously by people. We have also used combination of gestures such as simultaneous opening of mouth and squinting of eyes to toggle the mode between the cursor and scroll modes. Combination of gestures which are loosely connected prevents the false positive triggering of actions. Also the user mental memory has been considered while mapping the gestures to the actions. Certain actions such as left click and right click have been mapped to gestures like left and right wink. The purpose of this is to make the application very simple and intuitive to the users irrespective of whether they are novice, intermediate or experienced users. The other actions such as scrolling and cursor movement just require movement of the head/ looking in the desired direction just so as to displace the anchor point out of the box in the required direction.

D. Public Transport Assistance

To guide tourists and travellers in railway stations and bus stops, Ýatri Guideápplication is proposed. The application is targeted for all types of users irrespective of their physical or mental status. The motivation for the application comes from the fact that not all people may be in a position to physically operate public interfaces present in railway/metro stations. This can be due to the following reasons:-

- motor disability
- the hands being preoccupied such as due to carrying luggage
- hygiene concerns

Through this application, the user can check for express as well as local train schedules along with the bus schedules in the region of Maharashtra (MSRTC). The user needs to select the required source and destination stations and can see the available transport schedule on that route. The user also has an option to check the ticket fare on that route. Easy navigation between pages is supported in the application.

V. Human Computer Interaction Principles Used

A. Shneiderman's Eight Golden Rules

- Strive for consistency: The interface of our application is visually consistent across screen. There's also consistent mapping for directions and clicks like right wink for right click or look up to scroll up in scroll mode.
- Offer informative feedback: Our system provides a live feed of the user's facial expressions along with instructions to help the user perform the actions they require. The actions performed by the user is visually displayed to avoid confusion and dismiss any doubts regarding the action performed.
- Design dialogue to yield closure: The person directly in front of the interface is the one in control. Therefore distractions like people walking around the interface have no effect on the system. Other alternative accessibility controls like voice cannot distinguish such distractions in a noise environment. Therefore our system provides the user with a sense of complete control.
- Reduce short-term memory load: The number of actions limited to keep the system simple and easy to learn. All the movements are intuitively mapped to head movements. Additionally, instructions of how to control the system are visually displayed to easy memory load. This makes our system very user friendly.

B. Norman's 7 Principles

- Make things visible: Our interface have simple controls and it's very easy to user to learn. All the actions are visually listed to make it easier to discover them.
- Get the mappings right: The cursor moves in the same direction as that of the head in the horizontal and vertical plane. Left and right click are mapped to left and right wink. The mapping makes it easy for the user to remember and perform the actions while also reducing short-term memory load.

VI. Experimental Results and Analysis

A. The Interface:

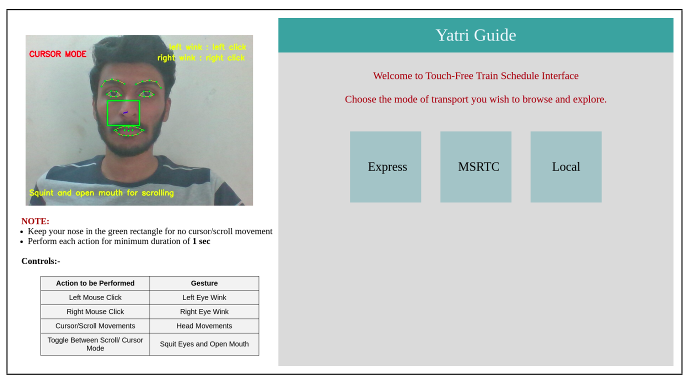

Fig. 3. Home Page for selecting mode of transport

Fig. 3 depicts the home page of the interface. The following assumptions were considered while designing the interface:-

- The interface is placed in the stations such that only the user will be in front of the camera. This is to prevent multiple faces being detected.
- A plain white background is present behind the user so that the on-frame instructions are clearly visible.
- The interface camera is at user eye-level.
- Interface camera is of high quality so as to minimize latency.

The left half of the screen is static for the entire application. Instructions valid for all pages are shown in the form of a table which can be referred by the users at any stage of using the interface. Other specific and frequently used instructions are also shown dynamically on the live video frame. Sufficiently large buttons have been provided for easy access to select the mode of transport.

After selecting the mode of transport, the user needs to enter the source and destination as shown in Fig. 4.

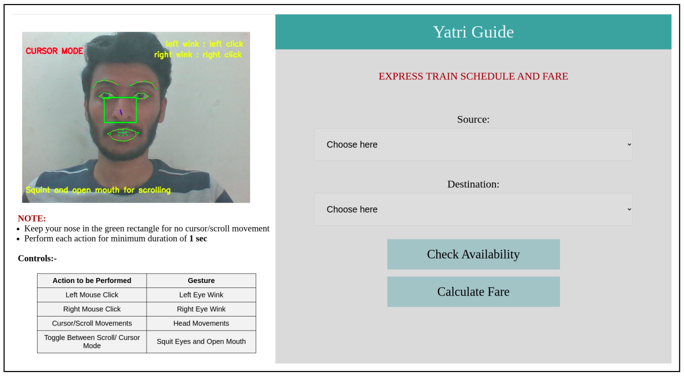

Fig. 4. Selecting source and destination

Fig. 5 shows the availability and fares for the source and destination selected in the previous page (Fig. 4). The user also has an option to check availability/fares for other routes directly using a home button.

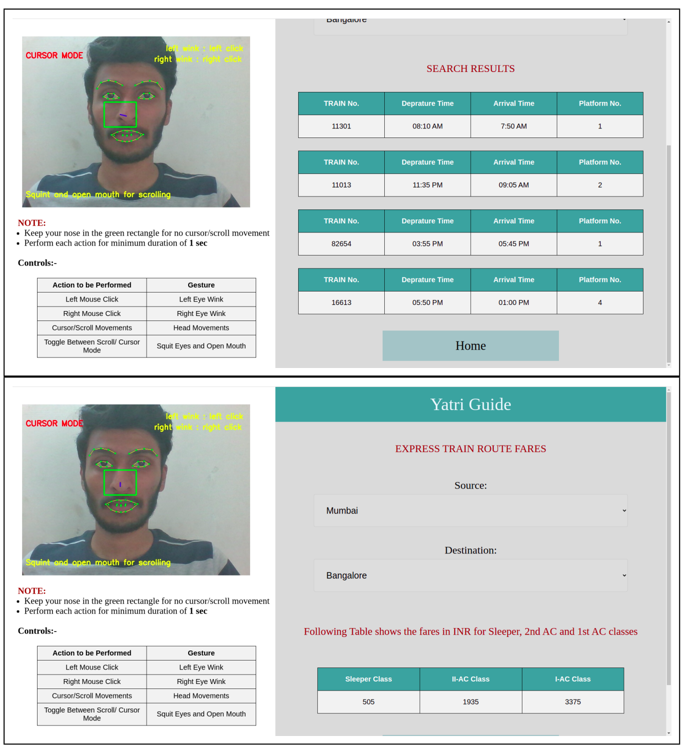

Fig. 5. Displaying search results / ticket fares on route

B. Empirical Analysis:

Empirical Analysis was performed for choosing the optimal threshold values for variables like horizontal cursor drag, vertical cursor drag and scroll drag. These three experiments were performed by novice and experienced users to support optimal and accurate use of the system yielding enhanced results. The experiment included participation by 25 users in total. the average time taken and the average number of misses was recorded for finding the optimal drag values.

1) Horizontal Cursor Drag: For choosing the optimal horizontal cursor drag, an experiment was performed such that it is independent of the vertical and scroll drag values. Table 2 shows the time required by the cursor to travel a distance of 1100 pixels on the screen, horizontally, to reach the target. We observe that the time taken to reach the target decreases and then increases with increase in the drag value. Additionally the number of times the target gets missed due to overshooting also increases with increase in the drag value. The drag value of 25 was chosen to be the optimal horizontal drag as it was the largest value which took the least amount of time to reach the target with no misses.

TABLE II
TIME TAKEN TO REACH TARGET WITH DIFFERENT HORIZONTAL CURSOR DRAG VALUE

Drag Value	Time (in sec)	No. of misses
5	29.58	0
15	11.22	0
25	7.72	0
35	25.11	2
45	43.32	5

2) Vertical Cursor Drag: Similarly, the optimal vertical drag was experimented and chosen such that it is independent of the horizontal drag and the scroll drag values. Table 3 shows the time required by the cursor to travel a distance of 600 pixels vertically to reach the target. It is observed that the time taken to reach the target decreases and then increases with increase in the drag value. Number of target misses also follow the same pattern as observed in the earlier experiment. Finally, 15 was chosen as the optimal drag value for vertical cursor movement even though the time required was more than that in the case of drag value 20 because the number of target misses in 15 drag was 0.

3) Scroll drag: The third empirical analysis involved in choosing the optimal scroll drag. Here, the chosen experiment did not depend on the horizontal and vertical cursor drags. Scroll drag values are smaller as compared to other drags as scroll drag is more sensitive to its movement. Table 4 shows the tested values for scroll drag. The time taken by the cursor to travel a distance of 2450 pixels vertically on an intermediate position on the screen is tabulated. The scroll drag of 3 is chosen as the optimal scroll drag value in this case.

TABLE III
TIME TAKEN TO REACH TARGET WITH DIFFERENT VERTICAL CURSOR DRAG VALUE

Drag Value	Time (in sec)	No. of misses
5	17.34	0
10	10.06	0
15	7.11	0
20	5.34	1
25	5.92	1
30	6.32	2
35	9.51	3

TABLE IV
TIME TAKEN TO REACH TARGET WITH DIFFERENT SCROLL DRAG VALUE

Drag Value	Time (in sec)	No. of misses
1	8.94	0
2	6.39	0
3	4.57	0
4	7.23	1
5	8.04	1
6	15.33	5

It is also observed that the vertical drag value is smaller than the horizontal drag value. This is because the width of the interface is larger than its height. This allows the buttons to be rectangular with a larger width. This makes it easier to navigate to buttons horizontally than vertically and hence, the horizontal cursor drag value is greater than the vertical cursor drag value.

C. Testing Procedure:

A fixed task was provided for all the users and the time taken to complete the task was averaged across several trials. The task consisted of finding the fares of express trains from Mumbai to Bangalore. The users were then asked to rate their experience on a scale of 1 to 5, 1 being the lowest and 5 being the highest rating. Table V tabulates the results obtained from the users.

D. Learnability Analysis:

We say that learnability is high if fewer tries are taken for novice users to reach the same amount of time required by an experienced user to perform the same task. Users were made to perform the same task multiple times and the average time taken in each iteration across all users was recorded for novice as well as experienced users. The results obtained are shown in Table VI.

TABLE V
NAVIGATION TIME TAKEN BY A USER TO PERFORM A TASK

User No.	Average Time (in sec)	User Rating
1	79.4	5
2	83.6	4
3	94.4	4
4	72.8	5
5	76.7	5
6	84.9	4
7	80.3	3
8	76.3	4
9	77.6	4
10	71.2	5
11	87.2	3
12	75.5	4
13	82.8	4
14	82.1	4
15	89.9	3
16	71.5	4
17	72.3	5
18	79.0	4
19	74.4	5
20	83.7	3
21	74.8	5
22	83.4	3
23	90.5	2
24	81.3	4
25	73.3	5

E. ANOVA Test

Analysis of Variance (ANOVA) test was performed for two versions of the built GUI. ANOVA test is used to check if there is any significant difference between two or more groups based on their means. Here, the 2 GUI versions are considered as

Task ID	Task Description	Task Duration (days)	Start Date	End Date	21/01/2022	27/01/2022	30/01/2022	31/01/2022	05/02/2022	06/02/2022	17/02/2022	09/03/2022	10/03/2022	07/03/2022	12/03/2022	13/03/2022	19/03/2022	20/03/2022	21/03/2022	10/04/2022
1	Detect facial features using Dlib shape predictor	7	21/01/2022	27/01/2022																
2	Map facial expressions to different actions	10	21/01/2022	30/01/2022																
3	Build cursor control system to detect facial expressions	10	27/01/2022	05/02/2022																
4	Select action to be performed using cursor control system	6	31/01/2022	05/02/2022																
5	Integrate the cursor control system	27	06/02/2022	09/03/2022																
6	Website (Front Page)	17	17/02/2022	10/03/2022																
7	Website (Check Availability)	7	07/03/2022	13/03/2022																
8	Website (Check Fares)	8	12/03/2022	19/03/2022																
9	Evaluation using HCI guidelines	9	13/03/2022	21/03/2023																
10	Analysis and Report	9	20/03/2022	10/04/2022																

Fig. 6. Gantt Chart

TABLE VI
LEARNABILITY ANALYSIS

Trial No.	Average Time for novice users (in sec)	Average Time for experienced users (in sec)
1	117	65
2	98	57
3	76	45
4	65	46
5	58	40
6	49	39

two groups and time is chosen as the independent variable. 10 users were tested upon for the ANOVA test. The average time taken across many trials by the users is recorded in Table VII.

Version 2 of our UI featured bigger buttons, a table with all instructions, displayed dynamic instruction based on current state, better organization and presentation of information on the screen by utilizing various colors. These additions made sure that the user had a clear idea of the actions to be performed to achieve the desired task and was never left at a loss of instructions. The result was a reduction in time to accomplish the task this is clearly visible in table VII

The ANOVA test summary is seen in Table. VIII. The significance level is set to 0.05, the degrees of freedom in the numerator is 1 which is calculated as one less than the number of groups and the degrees of freedom in the denominator is 8 which is calculated as the difference of the number of subjects and the number of groups. Hence, the corresponding F-critical value is 5.32. The obtained F-statistic value as shown in the table is 5.4712 which is greater than the F-critical value which in turn states that we must reject the null hypothesis. This

TABLE VII
TIME TAKEN TO REACH TARGET ON DIFFERENT GUI VERSIONS

GUI Version 1	GUI Version 2
85.3	71.2
80.7	77.6
78.4	72.8
81.9	83.6
83.1	75.5

is also proved using the p-value which is smaller than the significance level.

TABLE VIII
ANOVA SUMMARY

Source	DF	SS	MS	F-stat	P-value
Between Groups	1	82.369	82.369	5.4712	0.0475
Within Groups	8	120.4403	15.055	-	-
Total	9	202.8093	-	-	-

VII. CONCLUSION

In this research, we proposed and tested Yatri Guide, a novel touch-free interface for public transport assistance. The goal was to provide enhanced accessibility in public transport stations so as provide travellers with easy access to data. We achieved this with the help of a cursor control interface operated by mapping facial gestures to actions. Exhaustive empirical analysis was carried out to fix the cursor and scroll drag values so as to provide maximum user satisfaction. The testing procedure then consisted of assigning a fixed task to twenty-five users and measuring their time of completion. Each user were given multiple trials to determine the learnability of the interface. It was found that the time taken to successfully complete the task decreased with more number of trials proving the interface has a good learning curve. Further, the user rating of 4.04 on a scale of 5 showed the satisfaction levels of the users with Yatri Guide interface.

Individual contribution

Fig. 6 shows the gantt chart whereas Fig, 7 depicts the individual contribution.

Work Done	Suyash	Navaneeth	Naveen
Map facial movements to actions	✓	✓	
Getting input data and selecting action		✓	✓
Performing the selected action	✓		✓
Stream live video using flask backend + Website Integration	✓	✓	✓
Website :: Front Page		✓	
Website :: Check Availability	✓		
Website :: Check Fares			✓
Propose and Perform Analysis (metrics)	✓	✓	✓

Fig. 7. Individual Contribution

Base Paper

K. Sancheti, S. K. K, S. A and S. P, "Hands-free Cursor Control using Intuitive Head Movements and Cheek Muscle Twitches," TENCON 2018 - 2018 IEEE Region 10 Conference, 2018, pp. 0356-0361, doi: 10.1109/TENCON.2018.8650532.

References

[1] Khushal Sancheti et al. "Hands-free Cursor Control using Intuitive Head Movements and Cheek Muscle Twitches". In: *TENCON 2018 - 2018 IEEE Region 10 Conference*. 2018, pp. 0356–0361. DOI: 10.1109/TENCON.2018.8650532.

[2] Sarah M. O'Meara et al. "Comparing Two Different Cursor Control Methods which Use Single-Site Surface Electromyography". In: *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. 2019, pp. 1163–1166. DOI: 10.1109/NER.2019.8716903.

[3] Philipp Kopp et al. "Analysis and improvement of facial landmark detection". In: *Swiss Federal Institute of Technology Zurich, Zurich* (2019).

[4] Horatiu-Stefan Grif and Trian Turc. "Human hand gesture based system for mouse cursor control". In: *Procedia Manufacturing* 22 (2018), pp. 1038–1042.

[5] Sharon Mathew et al. "Eye Movement Based Cursor Control and Home Automation for Disabled People". In: *2019 International Conference on Communication and Electronics Systems (ICCES)*. IEEE. 2019, pp. 1422–1426.

[6] Davis E King. "Dlib-ml: A machine learning toolkit". In: *The Journal of Machine Learning Research* 10 (2009), pp. 1755–1758.