



DEFINING  
**GESTURES**  
FOR BASIC  
**WORD**  
**PROCESSING**  
TASKS IN  
**WEARABLE**  
**AUGMENTED**  
**REALITY**

PROJECT REPORT



IDC School of Design  
IIT, Bombay

Guide:

**Prof. Jayesh Pillai**

Submitted By:

**Kuldeep Singh Rathod**  
(176330001)

M.Des (Interaction Design)

# ACKNOWLEDGEMENT

I am sincerely grateful to my guide Prof. Jayesh Pillai, who with patience has guided me through the progress of the project, giving me the freedom to explore possibilities with this project.

I would also like to extend my sincere gratitude to the IxD faculty- Prof Venkatesh Rajamanickam, Prof Girish Dalvi, Prof. Ravi Poovaiah, Prof. Pramod Khambete, Prof. Vivek Kant and Prof. Anirudha Joshi for the constant feedback and constructive criticism that aided in my journey.

I thank my classmates for their valuable feedback and constant support and my family for their love and faith.

**Kuldeep Singh Rathod**

**176330001**

**IDC School of Design**  
**Indian Institute of Technology, Bombay**

# DECLARATION

I declare that this written document represents my ideas in my own words and where other' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any ideas/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the source which have thus not been properly cited or from whom proper permission has not been taken when needed.

**Kuldeep Singh Rathod**  
176330001

**IDC School of Design**  
**Indian Institute of Technology, Bombay**

# APPROVAL SHEET

The Interaction Design Project II entitled "**Defining gestures for basic word processing tasks in wearable augmented reality**" by **Kuldeep Singh Rathod**, roll number *176330001* is approved, in partial fulfilment of Masters in Design Degree in Interaction Design at the Industrial Design Centre, Indian Institute of Technology, Bombay.

Internal:

External:

Guide:

Chairperson:



# CONTENTS

## 01 01 ABSTRACT

## 03 02 INTRODUCTION

- 04 2.1 Types of AR
- 05 2.2 Interacting with AR devices

## 06 03 WHY WORD PROCESSING?

- 07 3.1 Operation Selection

## 08 04 BACKGROUND STUDY

- 08 4.1 Gestures
- 09 4.2 How to define gestures?
- 10 4.3 Gesture Classification
  - 10 4.3.1 Speech and gesture
  - 10 4.3.2 Screen based gestures
  - 11 4.3.3 Aigner's taxonomy

## 12 05 METHOD

- 12 5.1 User
- 13 5.2 Setup
  - 15 5.3.1 KSU (Keyboard Side User)
  - 16 5.3.2 GSU (Gesture Side User)
  - 17 5.3.3 Miscellaneous
- 18 5.3 Experiment Protocol
  - 18 5.3.1 KSU (Keyboard Side User)
  - 18 5.3.2 GSU (Gesture Side User)
  - 19 5.3.3 Controller

## 20 06 ANALYSIS

- 22 6.1 Data Collection
- 23 6.2 Results
  - 23 6.2.1 Successful Gestures
  - 25 6.2.2 All Gestures
  - 27 6.2.3 Overall Gesture Type Preference
  - 28 6.2.4 Ease of Defining a Gesture

## 29 07 LIMITATIONS

- 29 7.0.1 Unimodal Interaction
- 29 7.0.2 Compatibility amongst Participants

## 30 08 CONCLUSION

## 31 09 IMAGE REFERENCES

## 32 09 REFERENCES



# 01

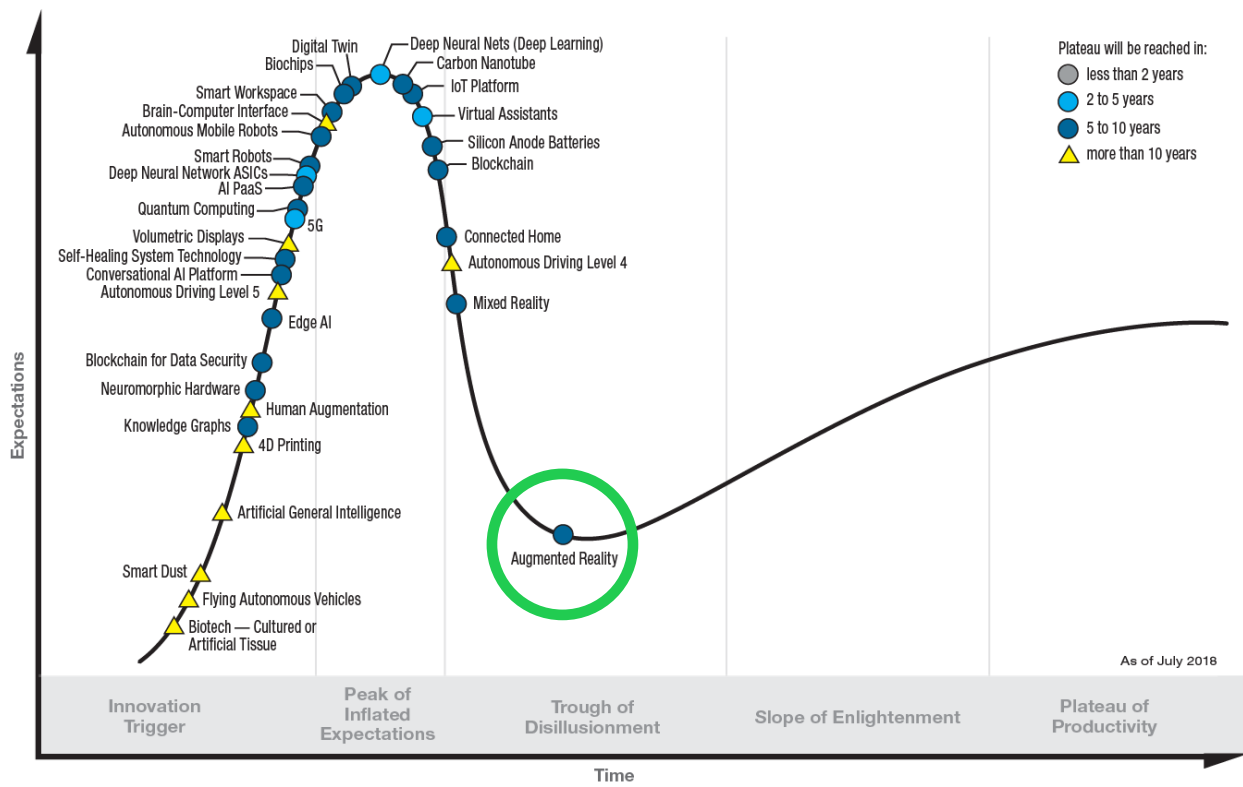
## ABSTRACT

Augmented reality has taken the reins from existing technology by integrating digital data in real time and providing an interactive experience for users. There have been a lot of research and technological advancements on hardware and software fronts to make these interactions easy and intrinsic. If we look at the wearable devices for augmented reality, hardware developments have not been able to keep pace with software developments. This acts as a limiting factor for the user experience of these devices.

Most of the basic activities of users on any technology devices includes tasks like internet surfing, reading and writing emails which require text input. The technology used by existing AR devices use point and click method which has been adopted from traditional GUI systems. This method to provide text input tends to be tedious and time consuming. This study has been conducted to identify and define the elements of appropriate gestures which will act as input for AR devices. The scope of the study is limited to word processing for specific

tasks chosen beforehand. The experiment undertaken for this study is in the form of a user elicitation study which is taken as the base to define guidelines for designing the specified gestures.

## Hype Cycle for Emerging Technologies, 2018



[gartner.com/SmarterWithGartner](https://gartner.com/SmarterWithGartner)

Source: Gartner (August 2018)  
© 2018 Gartner, Inc. and/or its affiliates. All rights reserved.



## MOTIVATION

According to Gartner Hype Cycle for 2018, Augmented Reality as well as Mixed Reality is in trough of disillusionment and moving towards slope of enlightenment (Fig 1.a). This brings us to the stage where we should think about the use of this technology in various spheres. There is a lot of room to work on usability of the AR devices and improve the user interactions. These interactions should be designed and standardised keeping user experience and usability in mind. This standardisation will provide an apt base for further research as well. Otherwise, it might lead to a situation where a lot of inappropriate practices are established and the technology is not utilized to its full potential.

Fig 1.a: Gartner Hype Cycle for Emerging Technologies, 2018

# 02

## INTRODUCTION

The technology of Digital Reality (AR, MR, VR) has been around for quite a few years now and it has the potential to offer solutions to a lot of problems which our existing devices (Laptops, Tabs, smartphones) have[1]. The focus has shifted from the product to users and the devices are being made to be more user friendly and convenient. If we talk in business context, a lot of industries have been very receptive of the new technologies and a variety of ecosystems have risen with the help of technology platforms.

The current applications of AR provides an enhanced user experience by overlaying digital information over the real world. Representing the next stage of technology evolution, it can revolutionize our lives the way smartphone did a decade ago. We are surrounded by devices like smartphones, laptops, tablets etc. but each one of the evolved devices has features that overcomes the limitations of the previous one. For example, Tablets are more portable than laptops and hence are more suitable for inspection tasks. Whereas smartphones takes

portability to another level but screen size is compromised. Similarly, Augmented Reality has the potential to provide solutions to all such problems simultaneously and thus it's quickly penetrating into various markets. It has full potential of becoming as ubiquitous as Smartphones are today.

# 2.1

## TYPES OF AR

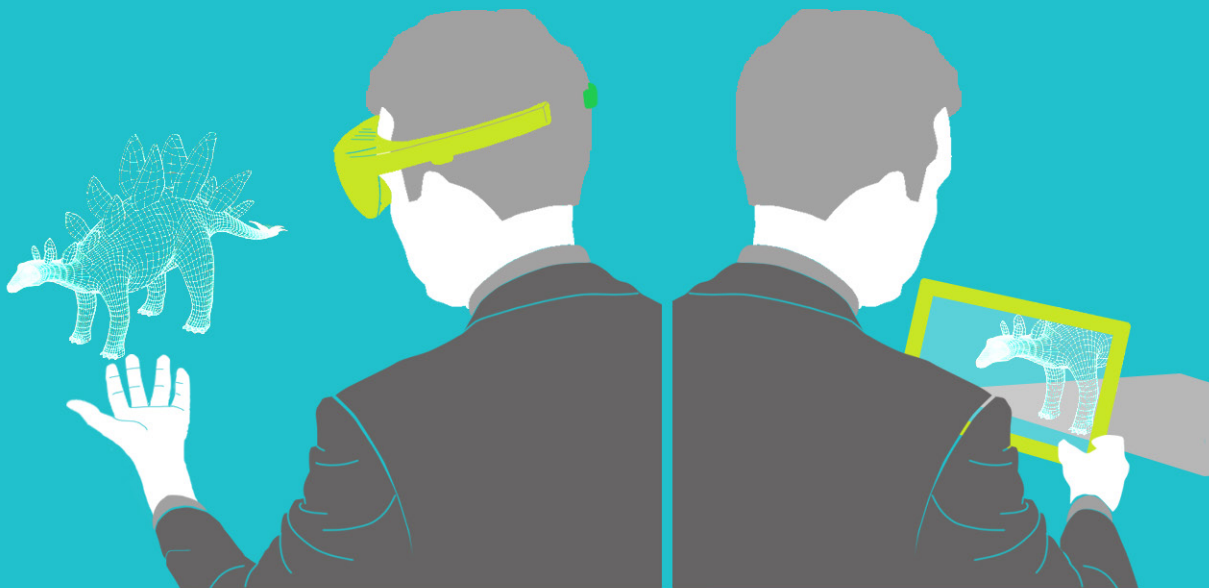


Fig 2.1.a: Types of Augmented Reality Experiences

Currently, there are two types of AR experiences available, screen based and glass based (Fig 2.1.a). Screen based AR refers to the AR applications available on tablets and smartphones which uses their existing hardware for creating augmentations. Interacting with these applications involves the use of touch based screen interactions which we normally use in our devices. The augmentations in this case are visible only through a screen. With the recent developments of technologies like ARCore and ARKit, We can establish relationships between the digital augmented content and the real world.

On the other hand, glass based (wearable) devices provide a more immersive experience by providing a better view with the digital content placed more appropriately in the real world. Although it has a limited field of view for now, but that's just a technological limitation at this point of time. With research and more advancements, these limitations will surely be rectified in the near future. The scope of incorporation of natural and intuitive interactions in this type of AR is much higher than screen based because of the physical freedom of both hands. With this, we can incorporate the modalities of gesture and speech for interaction with AR devices. Due to these reasons, this study has been focused on the wearable AR devices.

## 2.2 INTERACTING WITH AR DEVICES

In wearable AR devices, user interaction is the key to success of the device. There has been a lot of research to identify interaction methods with AR/VR devices. Mid-Air hand gestures being one of the most important one, since it can utilise the physical freedom of the user to interact with the device. Natural hand gestures provide a very convenient and intuitive interaction method which bridges both digital and the real world[5]. Hand gestures were primarily implemented as an add-on to the speech input in most of the studies based on multimodal AR interfaces[6], [7]. Though speech looks like a better solution, it has various limitations considering AR HMDs to be a potential ubiquitous technology. For example, imagine a public place filled with people giving audio commands to their devices or an office environment with people using voice commands in a meeting room. It will be very difficult to stop the interferences between the devices from people nearby, and more importantly there would be an issue of privacy (Fig 2.2.a).

Interactions in current wearable AR devices like HoloLens, Magic Leap, Meta 2 etc. are primarily dependent on hand gestures. But most of these gestures are a result of the technical limitations of the hardware. For example, the two primitive gestures in HoloLens, the Air tap and bloom gestures are designed to maximize detection, which again is a limitation caused by the current technology. Research also suggests that the gestures were designed by

researchers for optimal recognition rather than for naturalness, meaning that they were often arbitrary and unintuitive[8], [9]. This establishes the need for guidelines which can be used to design natural and intuitive gestures for AR devices before they become mainstream.

Although text and speech input represents a very vast area to work with and a lot of research has already been done on it. This study focuses on the interactions required after the key user input has been provided to the device. It majorly discusses the tasks which are used to process/manipulate the text i.e., word processing, since these gestures are very frequently used and demands more attention. The experiment has been conducted to study hand gestures for word processing tasks because they provide a more intuitive and immersive experience [10]-[12].



Fig 2.2.a: Issues with speech input

# 03

## WHY WORD PROCESSING?

HMDs like HoloLens and Daqri have already been adopted in various industrial sectors. This represents the first wave of market penetration, where the major players were targeted from manufacturing sector. The main adoption was from manufacturing sector due to the hardware developments and price range of the products, which was suited for this particular industry. Later on, these devices were developed keeping other professional needs in mind and adopted by other industries as well. For example, AR workspace by Microsoft which allows you to setup a virtual workspace with most frequently application arranged in 3D space around you. With increasing adoption of these devices which takes up a large share of our daily work, they will have frequent use of text input. Hence, increase in tasks of word processing.

The existing interfaces make use of our traditional QWERTY keyboard for text input (Fig. 3.0.a). People are trying different experiments to improve the text input in these devices. Some of them have been focused on hardware changes, while others

on the interaction part of the devices. This study focuses on the latter one since hardware dependency might compromise with the portability and the overall user experience. Assuming the advent of a more efficient AR based text input method, I will be discussing the ways to perform other three types of text related tasks: Manipulation, Formatting and Output. The most common mistakes committed by developers is to keep these options in the form menu buttons. This is the most outdated and unimaginative way of interacting with AR devices. With a new technology like this, We should not be forcing traditional WIMP GUI concepts. Work should be done in a new direction where all these tasks needs to be re-imagined and experimented.

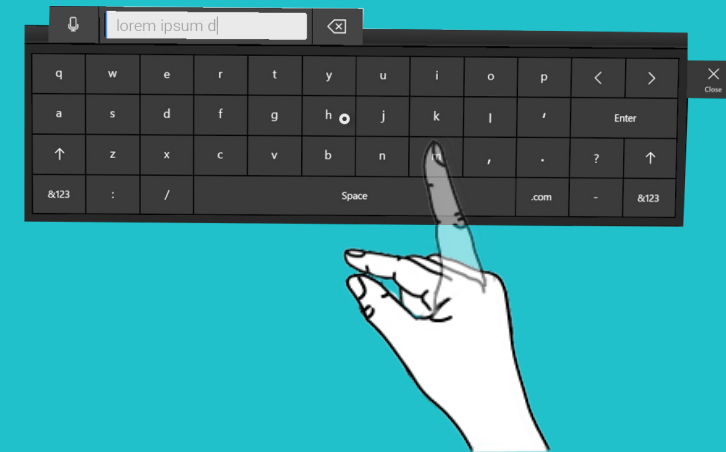


Fig 3.0.a: Text input in HoloLens

# 3.1

## OPERATION SELECTION

Word processing softwares have become much complex now and they can perform many more functions than text entry. But focus of this study has been on basic operations since they represents the majority of functions used. Overall, the basic word processing operations can be divided into five components: Text input, Text Formatting, Text manipulation and Output. Text input has not been covered in the scope of this study due to time constraints.

From a list of 16 basic operations, the 9 most frequently used ones were selected: Bold, Italics, Underline, Superscript/Subscript, Delete, Copy, Move and Paste. These operations are classified under word processing components and listed in the Fig: 3.1.a. During the experiment, these operations were arranged in a simple paragraph such that each operation appeared twice. This was done to certain the validity of a successful gesture as well as the ease of memorising it.

— — — — INPUT — — — — FORMATTING — — — — MANIPULATION — — — — OUTPUT — — — —

### FORMATTING

**Bold**  
**Italics**  
**Underline**  
**Subscript**  
**Superscript**

### MANIPULATION

**Cut**  
**Copy**  
**Paste**  
**Delete**

Fig 3.1.a: Selected Operations for the experiment

# 04

## BACKGROUND STUDY

### 4.1. GESTURES

Many novel interfaces like multi-touch, tangible, in-air gesture detection and organic user interfaces have tried to bring changes in the current WIMP type interactions. Since the development of touch screens, a lot of research has been done of gestures for devices ranging from mobile phones, large displays, consoles to AR devices [5], [13]–[16]. But the research for mid-air hand gestures has been very limited or have not been shared publicly to build on that work.

## 4.2 HOW TO DEFINE GESTURES?

There are two school of thoughts when it comes to defining gestures for a system: the first one suggests that gestures should be elicited. This means that it should be derived from potential users by putting them in their natural environment. This theory depends on the assumption that a new user's guess will produce the desired result[13], [17]. The second theory proposes that gestures should be designed after examining distinctive physical and psychological implications of a device in a given context, and then taught to users[18]–[20]. This approach can be used to design languages that not only avoid conflicts, but which are ergonomically feasible and have high bandwidth. For this study, the approach falls somewhere in the middle where (Fig: 4.2.a.) gestures were elicited from users. But instead of mapping these gestures directly to the set of operations, I have classified them to understand their psychological and physical structure. This classification can be used as a base by other designers to design more natural gestures.

The methodology followed in this study is quite similar to previously conducted elicitation studies, where participants were shown two situations and were asked to come up with a gesture to perform that manipulation[5], [13]. But to add a new element to this technique, I took inspiration from Aigner's experiment[21] and added a recognizer – a second participant, who observed the actions performed by the first participant via a muted video feed. I will

be referring to these participants as Gesture Side User (GSU/first participant) and Keyboard Side User (KSU/second participant). The task of GSU is to instruct certain tasks to KSU using gestures through a visual feed. So, in this case KSU acts as a feedback mechanism for GSU. This whole experiment imitated a unimodal gesture interface. The choice of using a human instead of a digital system was taken to prevent legacy bias[21]. This setup is quite similar to previous works of Aigner in CSCW, but the tasks, goals and information collected are different. In this study, the focus is on discrete tasks and no provisions have been given for continuous manipulation..

Fig 4.2.a: Methodology for defining gestures

## 4.3 GESTURE CLASSIFICATION

Gestures can act as a medium of input for several devices like mobiles, tablets, game consoles, tabletops, robots etc. The interaction techniques can be direct surface touch, mid-air gestures, full-body gestures or a combination of some or all[22]. Gestures can either be performed with the help of physical controllers(remotes, pens etc.) or using parts of human body. For this study, focus is on the latter i.e. free hand mid-air gestures.

### 4.3.1 Speech and gesture

Initial gestural classifications incorporated the relationship between gesture and speech [25][26]. Kendon[27] did the same thing by proposing classes of gesticulation (beat, cohesives), language-like (iconic), pantomimes, emblems (deictic), and sign language (symbolic). Then, he related them to accompanying speech by arranging them in a continuum of communication. Most of such works are modifications of Efron's[28] work. Similar works were further carried out by Cadoz[25] and McNeil[26]. Hence these classifications are not relevant for our study.

### 4.3.2 Screen based gestures

A lot of work has been done to define gesture for various other types of interfaces like, Wobbrock defined gestures for surface computing[13], Ruiz worked on motion gestures for mobile interac-

tions[13][14], Vatavu worked on free-hand TV control[15], Seyed on multi-display environments[16], and Piumsomboon on Augmented Reality[5]. Piumsomboon's study which is based on Wobbrock's taxonomy is completely based on guessability study. He classified gestures into form, nature, binding, flow, symmetry and locale, which only considers the physicality of the gesture performed. But a lot of gestures have cultural and social significance which may vary person to person. Moreover, the tasks performed during the study were discrete and lacked context. This is critical because gestural meaning can differ as per the context.

### 4.3.3 Aigner's Taxonomy

Word processing softwares have become much complex now and they can perform many more functions than text entry. But focus of this study has been on basic tasks since they represents the majority of functions used. Overall, the basic word processing tasks can be divided into five components: Text input, Text Formatting, Text manipulation and Output. Text input has not been covered in the scope of this study due to time constraints.

From a list of 16 basic operations, the 10 most frequently used ones were selected: Bold, Italics, Underline, Superscript/Subscript, Delete, Insert, Copy, Move and Paste. These tasks are classified under word processing components and listed in the Fig. 4.3.3.a. During the experiment, these tasks were arranged in a simple paragraph such that each task appeared twice. This was done to certain the validity of a successful gesture as well as the ease of memorising it.

#### Pointing

Gestures used to indicate objects or directions.

#### Semaphoric

Hand postures or movements that convey some meaning.

#### Pantomimic

Demonstration of a specific task performed without an object.

#### Iconic

Gestured used to communicate information about specific objects or entities.

#### Manipulation

Gestured are used to guide movement in a short feedback loop.

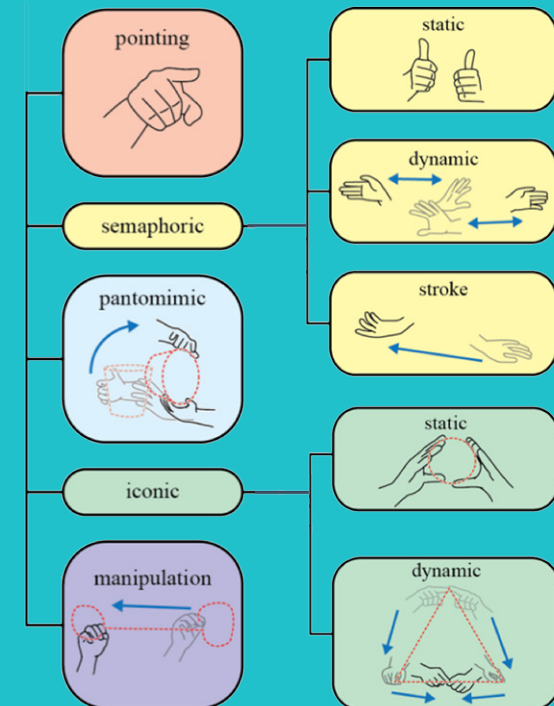


Fig 4.3.3.a: Aigner's Taxonomy

# 05 METHOD

## 5.1 USER

AR is still a new technology and very few people have used it and even lesser people have used devices like Hololens and Meta 2. In this study, users have been picked up to make sure that the participants had no prior experience of such wearables to avoid confounds caused by gestures learned from previous use of similar devices. It was also required that the users are technologically literate and well versed with the basics of at least one word processing software. Further, I have focused the experiment on young adults as previous researches suggest that age related changes in perceptual, cognitive and motor systems can act as confounds[25]. This allowed for minimising variability in user population. Age and genders were balanced in each group to minimize compatibility issues. The demographics of the experiments are presented below in Table 1. Left handed users were also included to check if preference of dominant hand had any effect on the gesture performed.

## 5.2 SETUP

The hardware used for the experiment consists of a mac mini(2014) and 3rd gen Microsoft Hololens. Both the devices were connected to the same wifi network (Fig: 5.2.a.). A php web server was setup on the laptop to allow data transfer between the devices . For recording the gestures performed by GSU two camera were set-up in GSU's room, capturing the video from side and front. A third recording was also done in the form of screen capture from KSU's interface to help sync the tasks done by both the participants. The layout of the experimental setup in both the rooms is shown in Fig 5.2.b, 5.2.c.

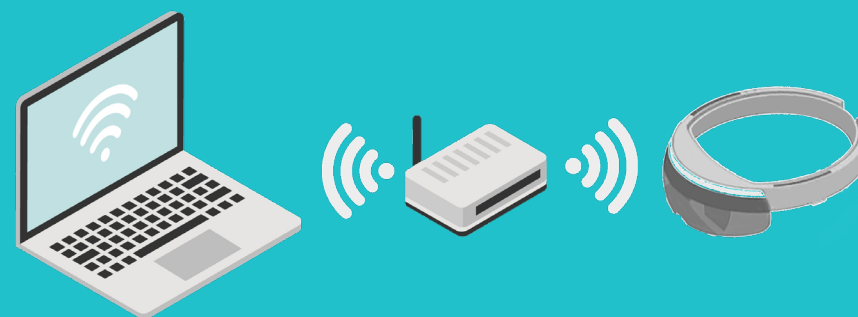


Fig 5.2.a: Network setup



Fig 5.2.b: Room setup

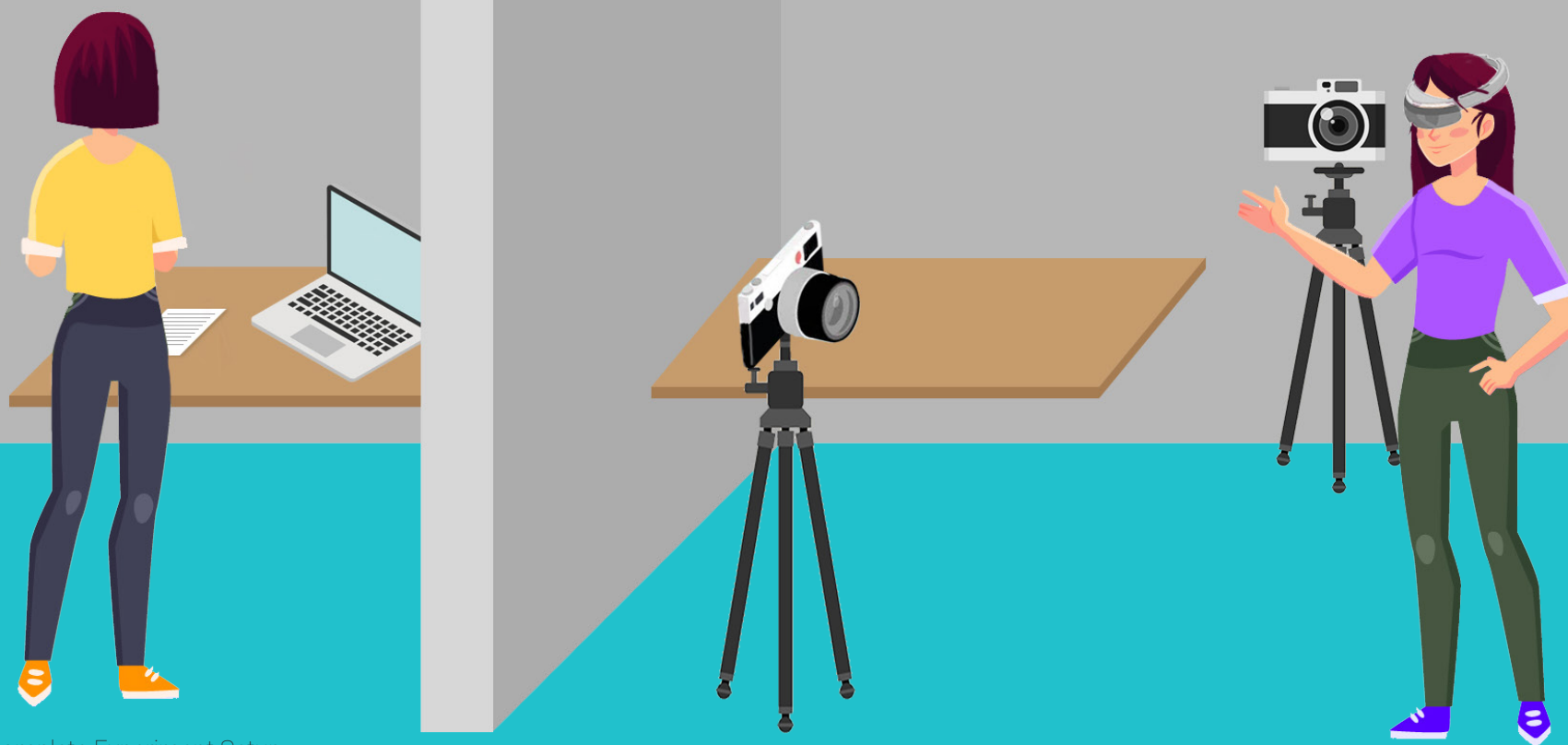


Fig 5.2.c: Complete Experiment Setup

### 5.2.1 KSU (Keyboard Side User)

The laptop's screen was divided into three parts: first part of the screen consisted of a web application running on google chrome, the second part had the live video stream from GSU's room and the third part showed the vision of GSU through HoloLens. The web application consisted of a text area and a tile strip on top containing buttons for accessing various word processing features. The interface of this app (Fig 5.2.1.a) was made quite similar to the existing word processing softwares to make it easier for KSU to work on it. Additionally, it has a toggle button below the text area which informs GSU to start the task. This button automatically turns off when KSU starts performing the required task on his screen. Participants will have five chances to perform each task, which will be marked by KSU via checkboxes below toggle button. After the completion of a task, she can move to the next task by clicking on the 'Next' button.

A task sheet (Fig: 5.2.3.a) was also handed over to KSU which had a paragraph composed of lines extracted from children's book. In between the words, some black boxes were placed. These black boxes were used as indicators for KSU to elicit gestures from the GSU.

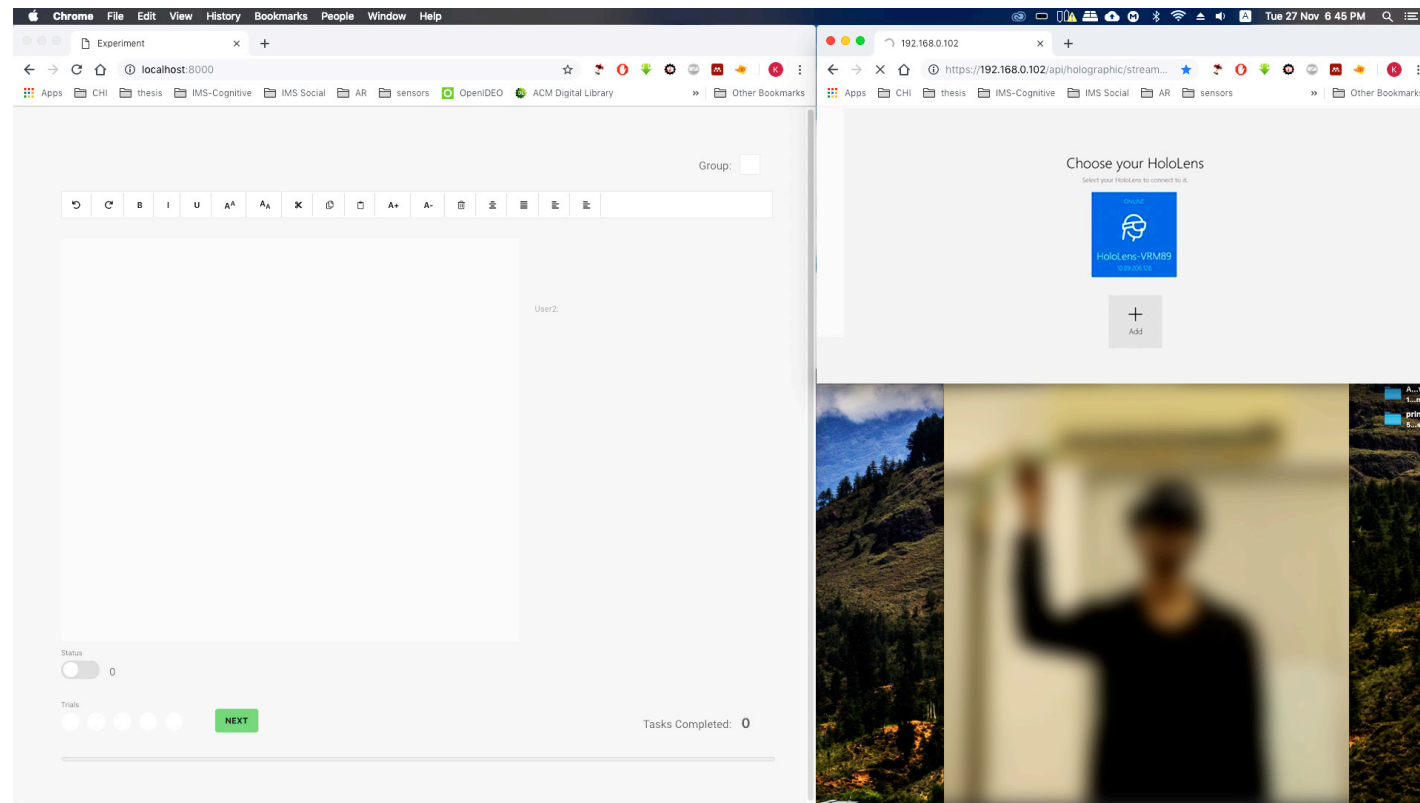


Fig 5.2.1.a: KSU's Interface

### 5.2.2 GSU (Gesture Side User)

The interface of the HoloLens application consists of a body-locked text area and a display-locked panel which has an indicator and a counter. When indicator turns green, GSU has to start performing the gesture. Counter depicts the number of trials performed for a specific task. GSU can see the text written by KSU in the web application and text edit functions used as well. (Fig: 5.2.2.a.)

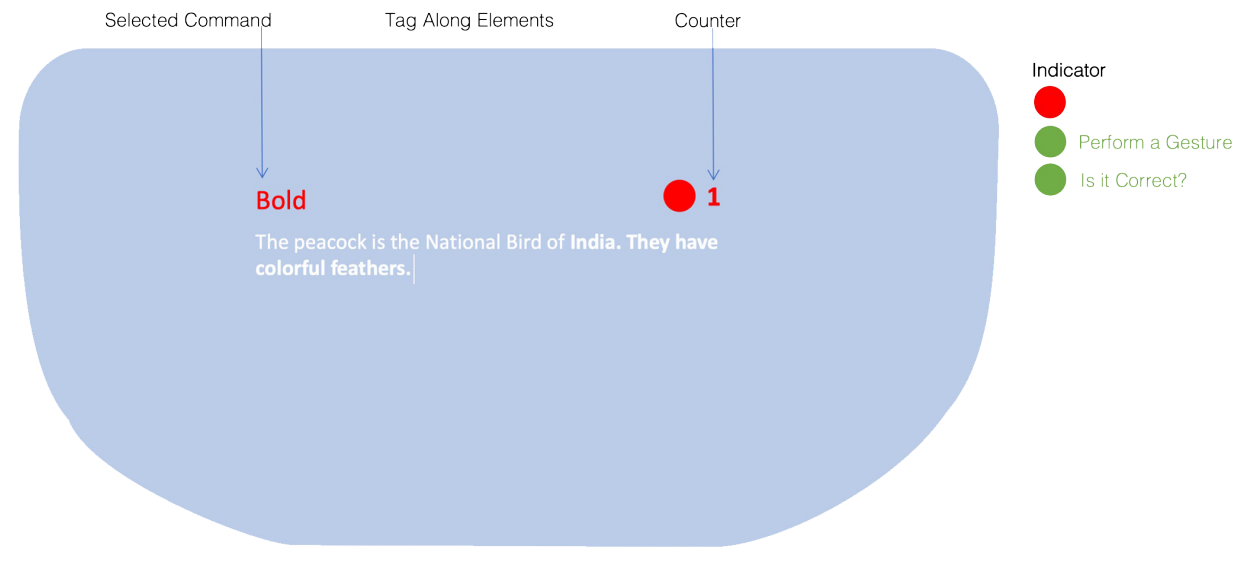


Fig 5.2.2.a: GSU's Interface

### 5.2.3 Miscellaneous

The controller of the experiment has placards, which have the tasks written on them derived from one operation each which were selected initially for the experiment. These placards (Fig: 5.2.3.b) are shown to the GSU once indicator is green. To remove any bias arising due to a particular order of these tasks, these placards are divided into three sets which randomises the order of tasks performed by each participant pair.

The peacock is the national bird of ■ India. They have colourful ■ feathers, two legs ■ and a small beak. They ■ are famous for their dance. When ■ a peacock dances ■ it spreads its feathers ■ like a fan. It has a long ■ shiny dark blue neck. ■ Peacocks are mostly found ■ in the fields. They are very beautiful ■ birds. The ■ females are known as Peahen. ■ Their feathers are used ■ for making jackets, purses etc. We can see them in a zoo.

Fig 5.2.3.a: Task Sheet

**Cut**  
**Previous word**

A:1 B:8 C:10

Fig 5.2.3.b: Task Placard

## 5.3 EXPERIMENT PROTOCOL

The participants were first divided into groups of two, uniformly distributing them on the basis of age and gender. Each group was given a common brief in the start, explaining about the experiment process and the final output expected from them. They were asked to come up with the signs for common understanding of approval, disapproval, selection of previous/next and completion of the task. This was done to help them communicate during the experiment. Then the participants were sent to separate rooms as KSU and GSU according to their preference.

Next, both the participants were individually informed about their role and setup of the experiment as explained.

### 5.3.1 KSU

She started the experiment by typing the text from the provided task sheet until the first black box. At every black box, she was supposed to turn on the toggle button indicating GSU to perform the gesture. Then she looked at the live stream to identify the gesture performed by GSU. After the gesture was complete, she had to increase the number on the counter by selecting the counter radio button. If she understood the gesture, she was supposed to perform the task depicted by the gesture and turn on the toggle button to get GSU's approval. If the task was done correctly, she could move on to the

next task by pressing 'Next' button. Otherwise, she had to undo the changes performed during the task and turn on the toggle button again. GSU would perform another gesture for the same task. For each task, a maximum of 5 trials were possible. After 5 trials are over, KSU was supposed to click on the 'Next' button to move on to the next task irrespective of getting the task right or not. The experiment went on until the last task was finished.

### 5.3.2 GSU

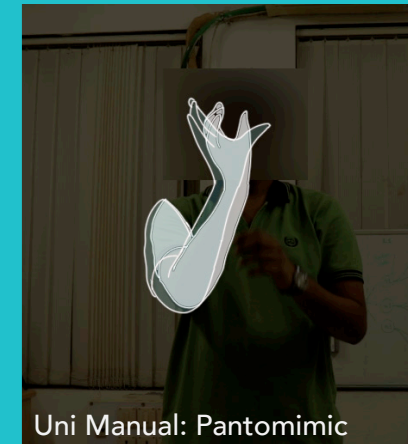
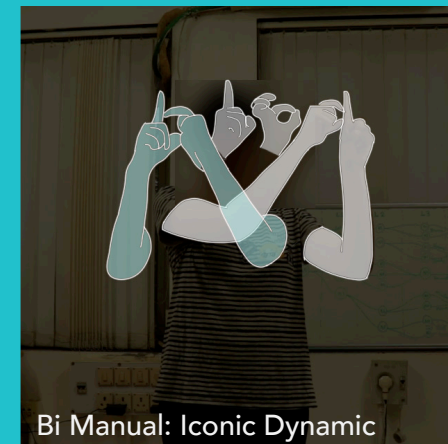
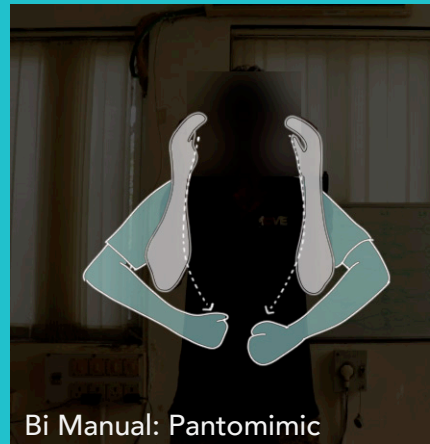
For GSU, the experiment started once the indicator turns green. She would have asked the controller to show the task placard. Then she was supposed to perform a gesture according to her understanding explaining the task to the KSU. After gesture was performed, she used the communication sign to signal the completion of the gesture. Once KSU started performing the task, it was visible to the GSU on her Hololens interface. KSU gave the confirmation of task completion by turning on toggle button and GSU's indicator will turn green. GSU gave her approval or disapproval to the KSU through the visual signals. If the task was done correctly, they were supposed to move on to the next task. Otherwise GSU had to perform another gesture for the same task after the indicator turns green. The process went on for 5 trials as explained earlier unless the controller informed GSU about the end of the experiment.



# 06 ANALYSIS

Few of the gestures performed by the participants are documented in Fig 6.0.a and Fig 6.0.b. The grey coloured part depicts the initial position of body part whereas sea green part depicts the final position of body part. The intermediate motion path is depicted through a dotted arrow line.

**BOLD**



**ITALICS**

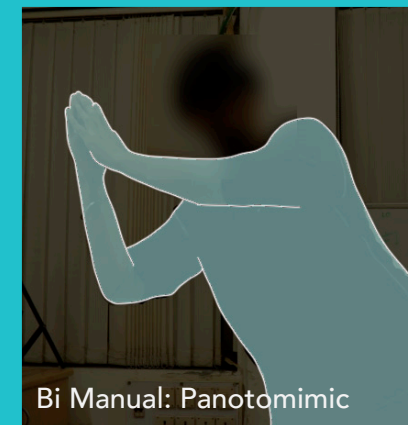
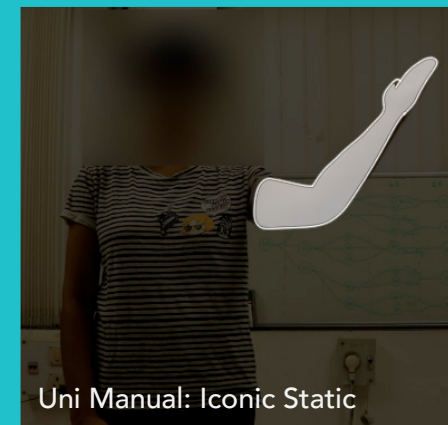
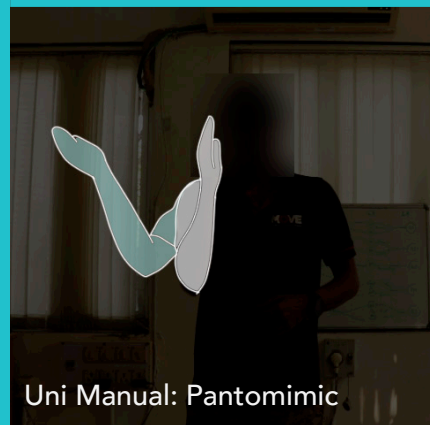
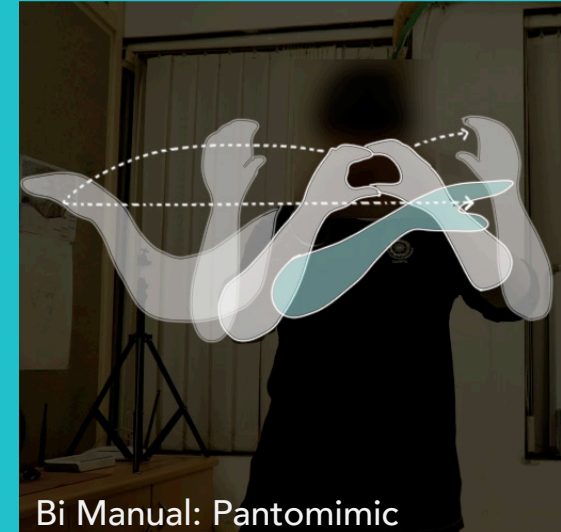
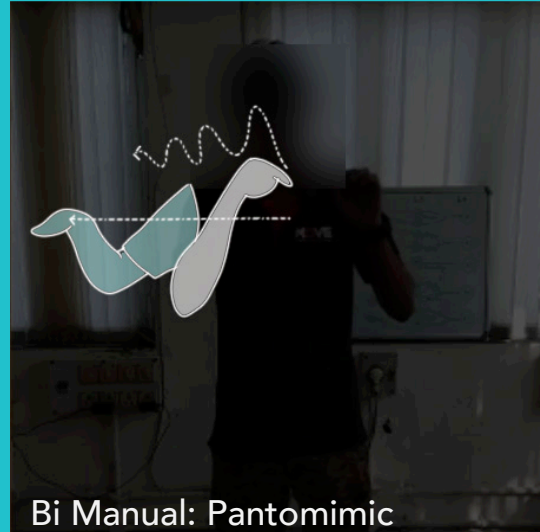
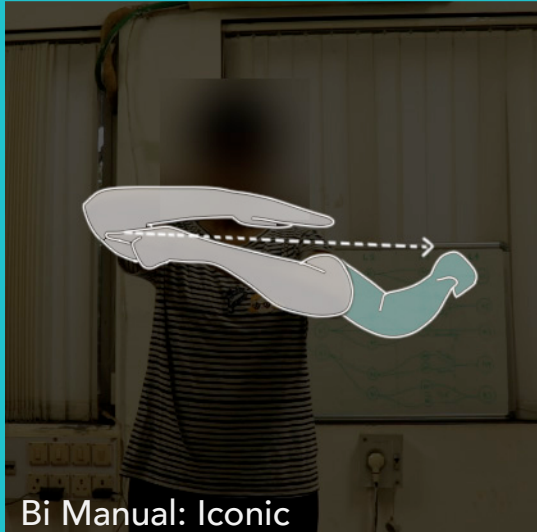


Fig 6.0.a: Few gestures performed for bold and italics

UNDERLINE



DELETE

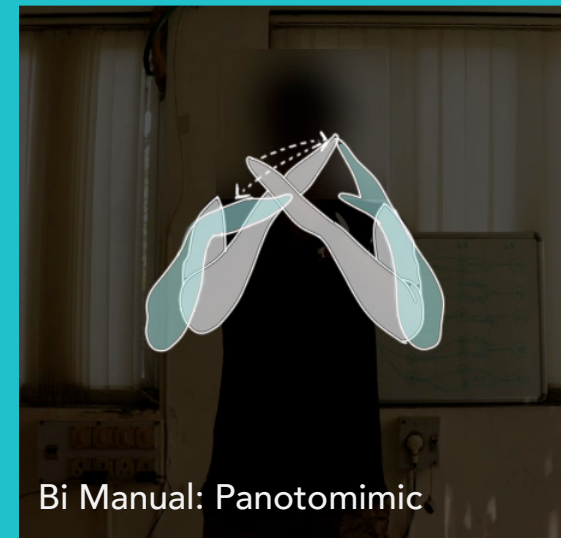
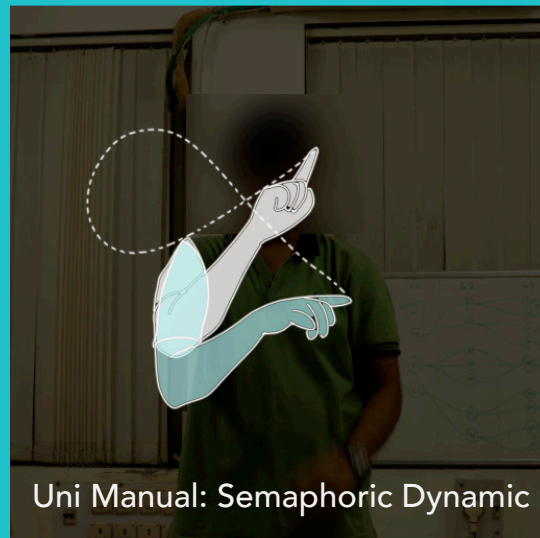
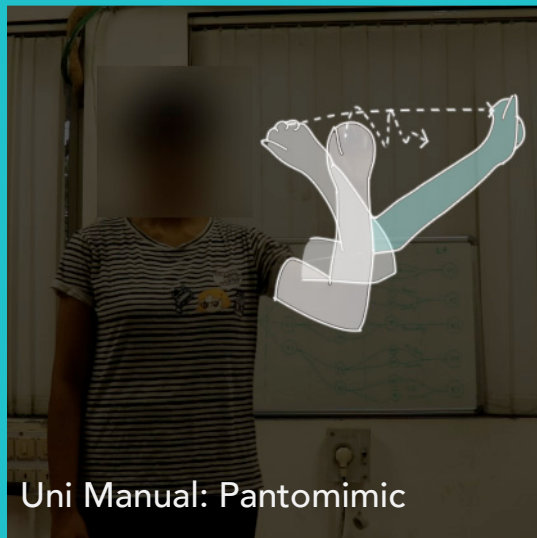


Fig 6.0.b: Few gestures performed for underline and delete

# 6.1 DATA COLLECTION

In the Data collection stage, the recorded footage from all the three cameras were compiled together group wise. For each group, it was further subdivided according to the task performed. Within each task, discrete trials were identified and recorded. All these details were recorded in a tabulated form and grouped operation wise along with their group numbers as shown in Fig 6.1.a. For example, the first green cell in the figure represents the first task performed by group 1 for the operation 'Bold'. The green color represents the successful trials and white represents unsuccessful trials.

Each trial was then closely observed to identify the gesture as shown in Fig 6.0.a & Fig 6.0.b. All the successful gestures were separated from unsuccessful ones and two sets were made. For each set, the number of hands used were identified. The complex gestures were broken down into individual parts. This helped in identifying the gestures which indicated the task and supporting gestures were neglected. Finally the gestures from both the sets were mapped with the Aigner's category and naming convention was set accordingly.

For example, cell named '2-Pa' represents the successful task performed in first attempt for the operation 'Bold' by group 1, which was done using 2 hands and mapped with the category Pantomimic in Aigner's taxonomy.

	Bold		Italic		Underline		Cut		Copy		Paste
Gp 1	2-Pa		1-Pa		2-ID		1-Pa		2-Pa		2-Pa
	2-Pa	Gp 1	1-Pa		2-ID		1-Pa		2-Pa		2-Pa
	2-Pa	Gp 4	2-Pa		2-ID		1-Pa		2-Pa		2-Pa
Gp 4	2-ID		2-IC		2-Pa		1-Pa		2-Pa		2-Pa
	2-ID		1-Pa		1-Pa		2-ID		2-Pa		2-Pa
	2-ID	Gp 2	1-Pa		2-ID		1-Pa		2-Pa		2-Pa
	2-ID		1-ID		2-ID		2-Pa		2-Pa		2-Pa
Gp 2	2-SC		1-ID		1-ID		2-Pa		2-Pa		2-Pa
	2-SC		1-Pa		2-Pa		2-Pa		2-Pa		2-Pa
	2-SC		2-Pa		1-ID		1-Pa		2-Pa		2-Pa
Gp 6	2-ID		1-Pa		1-ID				2-ID		2-Pa
	2-ID	Gp 6	1-Pa		1-Pa				2-ID		2-Pa
	1-ID		2-Pa		2-ID				2-Pa		2-Pa
Gp 7	1-ID		2-IC		2-ID				2-Pa		2-Pa
	1-ID		1-Pa		2-ID				2-Pa		2-Pa
Gp 3	2-Pa		1-Pa		2-ID				2-Pa		2-Pa
	2-Pa	Gp 3	1-IC		2-ID				2-Pa		2-Pa
	2-Pa		2-Pa						2-Pa		2-Pa
Gp 5	2-IC		1-Pa						2-Pa		2-Pa
	2-IC								2-Pa		2-Pa
	2-ID								2-Pa		2-Pa
	2-ID								2-Pa		2-Pa
Gp 3	2-ID								2-Pa		2-Pa
	2-ID								2-Pa		2-Pa
	2-IC								2-Pa		2-Pa
Gp 5	2-SC								2-Pa		2-Pa
	2-Pa								2-Pa		2-Pa
	2-Pa								2-Pa		2-Pa

Fig 6.1.a: Segregation of gestures in tabular form

# 6.2 RESULTS

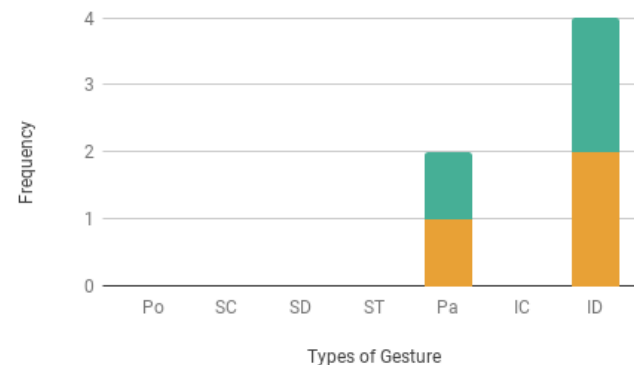
## 6.2.1 Successful Gestures

To obtain the results, a frequency distribution was plotted for all the successful gestures pertaining to every operation in the form of a stacked bar chart. The analysis was done separately for test manipulation and text formatting operations. It was observed that preference for bi-manual gesture was observed in text formatting operations (Fig 6.2.1.a), whereas in manipulation operations two very discrete behaviours were observed (Fig 6.2.1.b). For some manipulation operations like 'Cut' and 'Delete', participants preferred single hand, while for 'Copy' and 'Paste', participants used bi manual gestures. Highest variability in the types of gesture performed was observed for 'Bold' in Formatting operation whereas for 'Delete' in Manipulation operations. The most consistent gesture types were performed for 'Paste'.

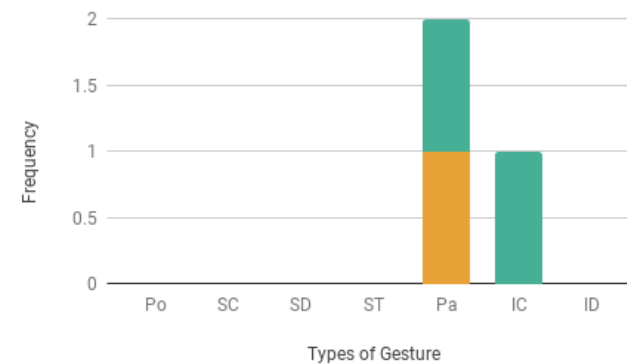
For successful text formatting operation, a preference for pantomimic and Iconic gestures was observed, whereas for successful manipulation operations all except for 'Delete' preferred pantomimic. For 'Delete' operation, Semaphoric Dynamic was preferred. When similar traits were observed in the total gesture plotting, a clean preference for pantomimic was observed in manipulation operations whereas the preference for Iconic Dynamic gestures takes over the preference for pantomimic gesture.

## Successful Gestures Text Formatting

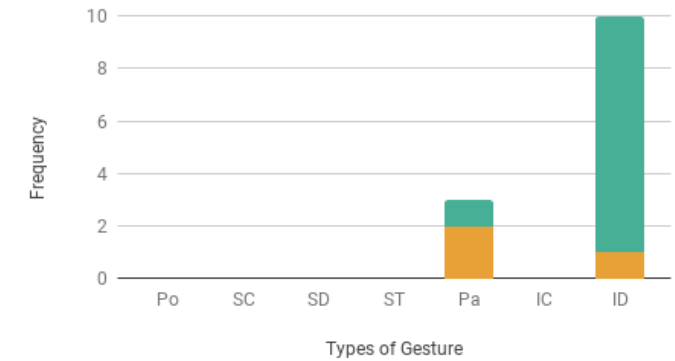
SUBSCRIPT (Successful Gestures)



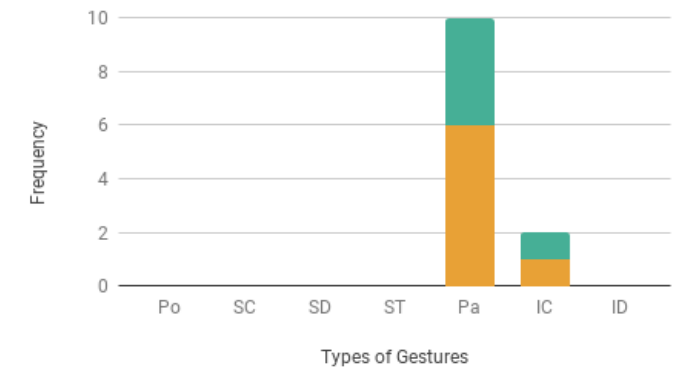
SUPERSCRIP (Successful Gestures)



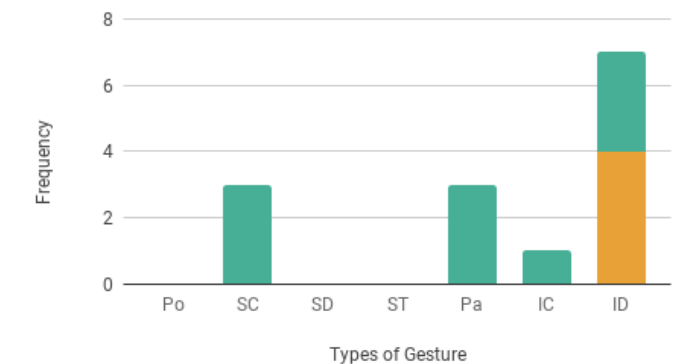
UNDERLINE (Successful Gestures)



ITALICS (Successful Gestures)



BOLD (Successful Gestures)



■ Bi-Manual ■ Uni-Manual

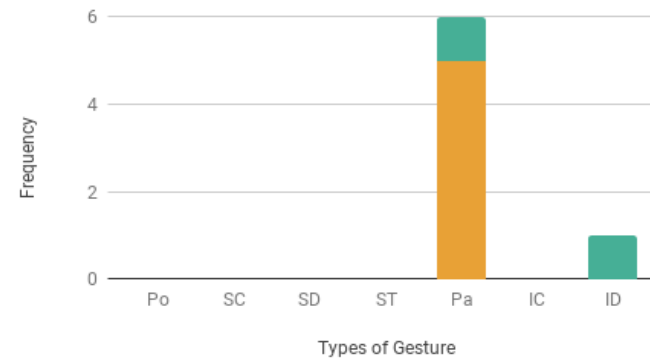
■ Bi-Manual ■ Uni-Manual

Fig 6.2.1.a: Frequency plots of successful text formatting operations

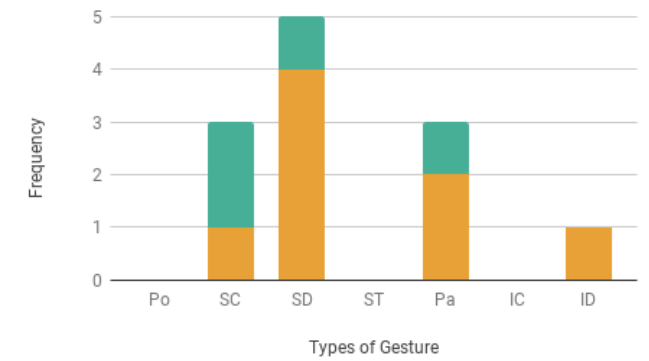
# Successful Gestures

## Text Manipulation

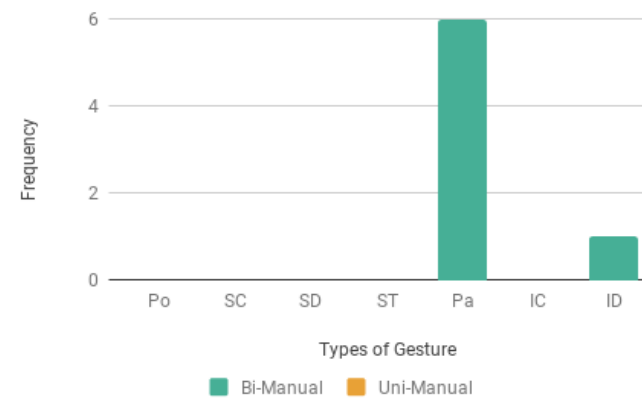
CUT (Successful Gestures)



DELETE (Successful Gestures)



COPY (Successful Gestures)



PASTE (Successful Gestures)

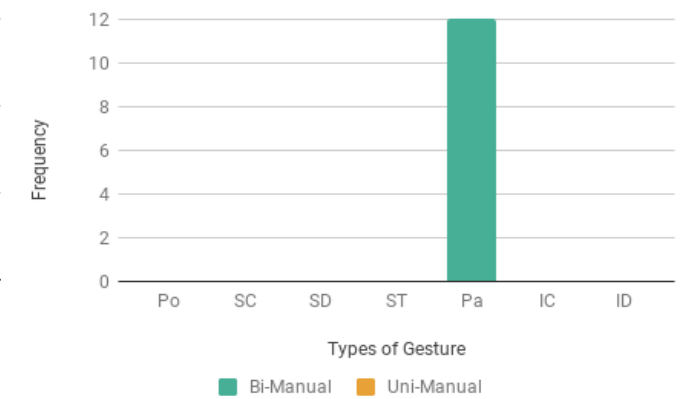


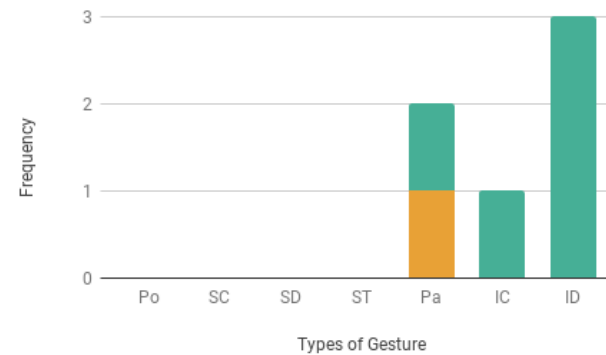
Fig 6.2.1.b: Frequency plots of successful text manipulation operations

### 6.2.2 All Gesture

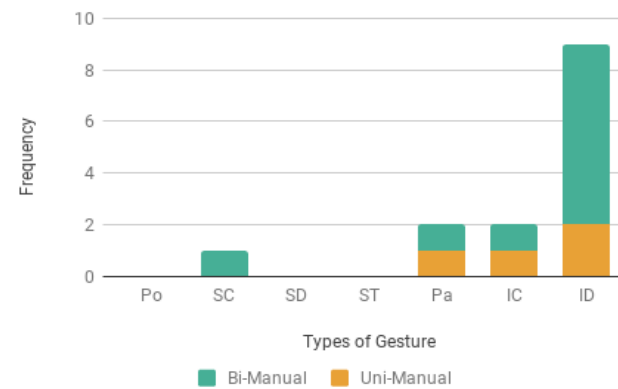
When all the performed gestures were plotted in the similar manner, the results almost remained same for manipulation operations whereas for formatting operations the variability for each operation increased (Fig 6.2.2.a and Fig 6.2.2.b).

## All Gestures Text Formatting

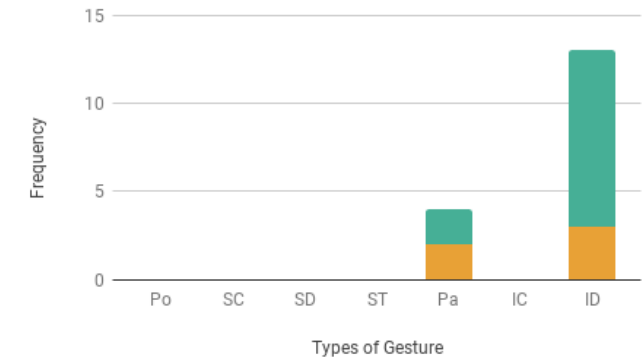
**SUPERSCRIP (All Gestures)**



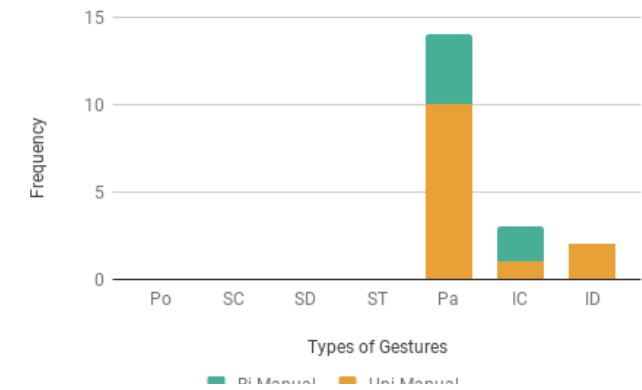
**SUBSCRIPT (All Gestures)**



**UNDERLINE (All Gestures)**



**ITALICS (All Gestures)**



**BOLD (All Gestures)**

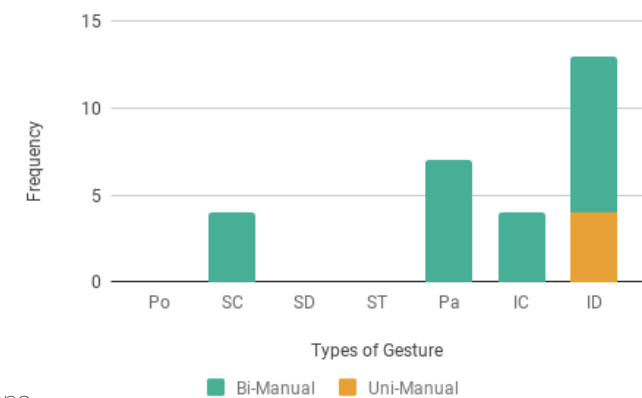
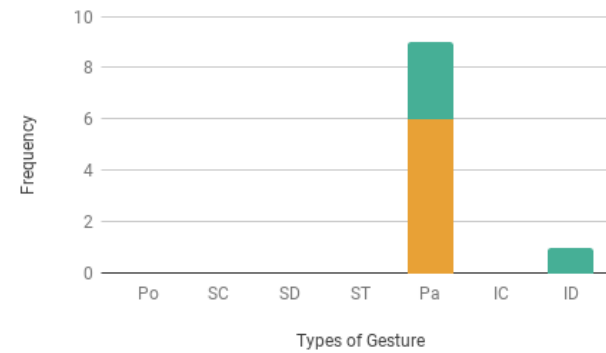


Fig 6.2.2.a: Frequency plots of All text formatting operations

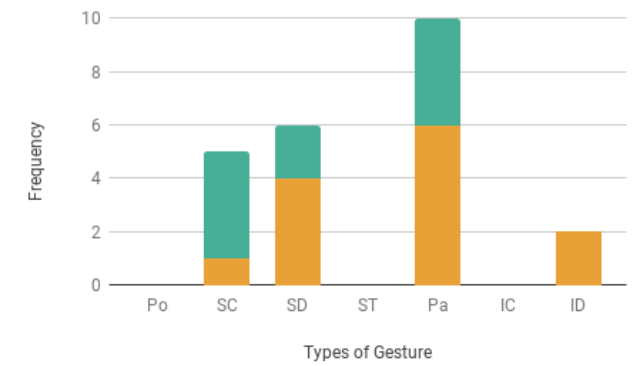
# All Gestures

## Text Manipulation

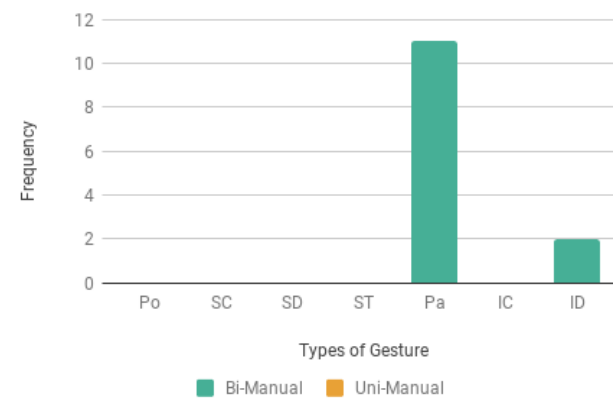
**CUT (All Gestures)**



**DELETE (All Gestures)**



**COPY (All Gestures)**



**PASTE (All Gestures)**

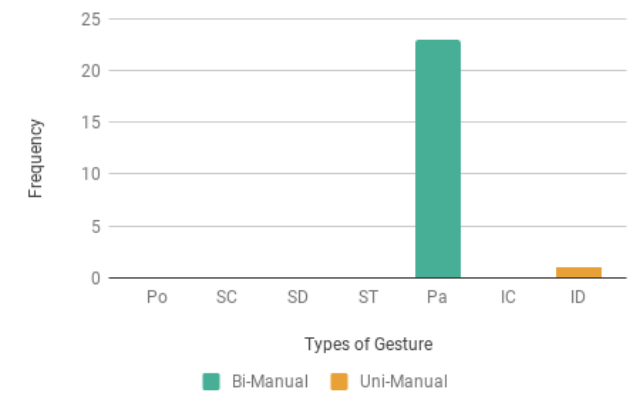


Fig 6.2.2.b: Frequency plots of All text manipulation operations

### 6.2.3 Overall Gesture Type Preference

When frequencies of all the performed gestures were plotted in a single chart a preference for bimanual pantomimic gesture was observed (Fig 6.2.3.a).

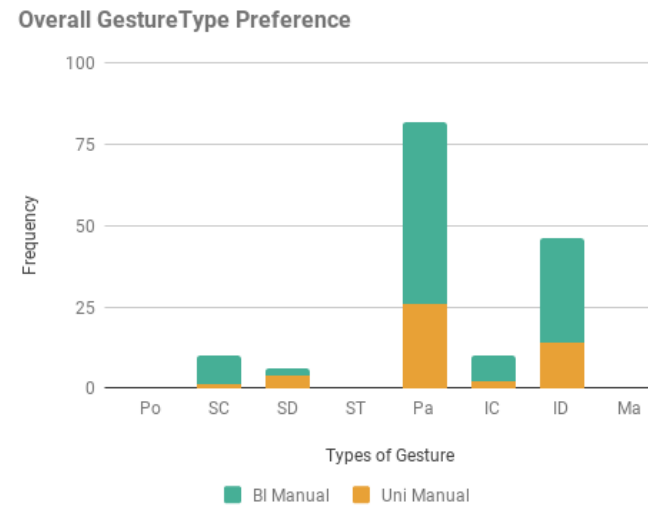


Fig 6.2.3.a: Frequency plots of All gestures

## 6.2.4 Ease of defining a gesture

For identifying the ease of defining a gesture for a particular operations, two plots were generated (Fig 6.2.4.a).

**First:** The frequency of all the operations being identified in the first trial was plotted against each operation. From this plot, most easily identifiable gestures were identified i.e., 'Underline' and 'Cut'.

**Second:** The highest frequency for each type of gestures performed for each operation was plotted. This was done to identify the operation for which most consistent gestures were performed i.e., 'Paste'.

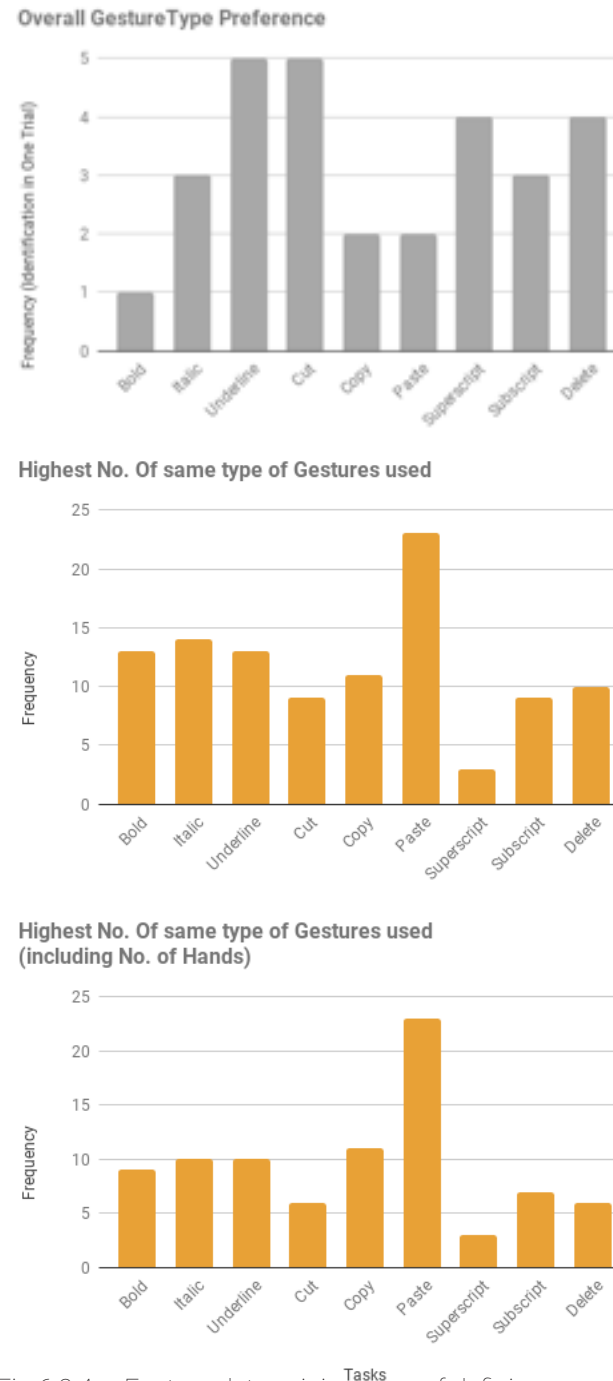


Fig 6.2.4.a: Factors determining ease of defining a gesture

# 07 LIMITATIONS

## 7.0.1 Unimodal Interaction

The entire study was conducted assuming the whole system to be unimodal which is not the case. In a multimodal interaction scenario, there will be effect of other modalities on the gesture input. So in actual cases, these effects should be taken into account.

## 7.0.2 Compatibility amongst Participants

To minimise the dependency on compatibility, the participants were recruited in groups instead of forming the groups later. This was done to make sure that the participants knew each other beforehand and had good compatibility with each other. The crux of this study was being able to elicit more natural gestures from the participants and Identification of gestures was only the secondary goal. Compatibility of participants might come into the picture for the identification of gestures alone. In the current setup, the compatibility of these participants cannot be quantified. Thus, the results might have some error due to this.

# 08

## CONCLUSIONS

The above theory is based on the pilot run carried out to define hand gestures for word processing tasks. This experiment is still work in progress. The final guidelines to classify and design gestures on basis of insights generated from this experiment will be shared later.

# 09 IMAGE REFERENCES

- Fig 1.a** : Gartner Hype Cycle for emerging technologies 2018  
(<https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018>)
- Fig 2.1.a** : Types of Augmented Reality Experiences  
(Author)
- Fig 2.2.a** : Issues with speech Input  
(Author)
- Fig 3.0.a** : Text Input in Hololens  
(Author),(graphics from Freepik.com)
- Fig 4.2.a** : Methodology for defining gestures  
(Author)
- Fig 4.3.3.a** : Aigner's Taxonomy  
(Roland Aigner)
- Fig 5.2.a** : Network Setup  
(Author),(graphics from Freepik.com)
- Fig 5.2.b** : Room Setup  
(Author),(graphics from Freepik.com)
- Fig 5.2.c** : Complete Experiment Setup  
(Author),(graphics from Freepik.com)
- Fig 5.2.1.a** : KSU's Interface  
(Author)
- Fig 5.2.2.a** : GSU's Interface  
(Author)
- Fig 5.2.3.a** : Task Sheet  
(Author)
- Fig 5.2.3.b** : Task Placard  
(Author)
- Fig 5.3.a** : Experiment Flow Diagram  
(Author)
- Fig 6.0.a** : Few Gestures performed for Bold and Italics  
(Author)
- Fig 6.0.b** : Few Gestures performed for Underline and Delete  
(Author)
- Fig 6.1.a** : Segregation of performed gestures in Tabular form.  
(Author)
- Fig 6.2.1.a** : Frequency plot of successful text formatting operations  
(Author)
- Fig 6.2.1.b** : Frequency plot of successful text manipulation operations  
(Author)
- Fig 6.2.2.a** : Frequency plot of all text formatting operations  
(Author)
- Fig 6.2.2.b** : Frequency plot of all text manipulation operations  
(Author)
- Fig 6.2.3.a** : Frequency plot of all gestures  
(Author)
- Fig 6.2.4.a** : Factors determining ease of defining a gesture.  
(Author)

# 09

## REFERENCES

- [1] W. Moberg, J. Pettersson, and V. Sundstedt, “DEGREE PROJECT FOR MASTER OF SCIENCE IN ENGINEERING GAME AND SOFTWARE ENGINEERING Interacting with Hand Gestures in Augmented Reality: a Typing Study,” 2017.lk
  
- [5] T. Piumsomboon, A. Clark, M. Billingham, and A. Cockburn, “User-defined gestures for augmented reality,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8118 LNCS, no. PART 2, pp. 282–299, 2013.
  
- [6] M. Kölsch, R. Bane, T. Höllerer, and M. Turk, “Multimodal interaction with a wearable augmented reality system,” *IEEE Comput. Graph. Appl.*, vol. 26, no. 3, pp. 62–71, 2006.
  
- [7] Heidemann, “Multimodal Interaction in an Augmented Reality Scenario,” *{Icmi} 2004*, pp. 53–60, 2004.
  
- [8] E. Kaiser et al., “Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality,” *Proc. 5th Int. Conf. Multimodal interfaces - ICMI '03*, p. 12, 2003.
  
- [9] T. Lee and T. Höllerer, “Handy AR: Markerless inspection of augmented reality objects using fingertip tracking,” *Proc. - Int. Symp. Wearable Comput. ISWC*, pp. 83–90, 2007.
  
- [10] M. Kim and J. Y. Lee, “Touch and hand gesture-based interactions for directly manipulating 3D virtual objects in mobile augmented reality,” *Multimed. Tools Appl.*, pp. 1–22, 2016.

- [11] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *AI Rev.*, vol. 43, no. 1, pp. 1–54, 2015.
- [12] R. O'Hagan and A. Zelinsky, "Visual gesture interfaces for virtual environments," *Proc. - 1st Australas. User Interface Conf. AUIC 2000*, vol. 14, pp. 73–80, 2000.
- [13] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, "User-defined gestures for surface computing," *Proc. 27th Int. Conf. Hum. factors Comput. Syst. - CHI 09*, p. 1083, 2009.
- [14] J. Ruiz, Y. Li, and E. Lank, "User-defined motion gestures for mobile interaction," *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*, p. 197, 2011.
- [15] R.-D. Vatavu, "User-defined gestures for free-hand TV control," *Proc. 10th Eur. Conf. Interact. tv video - EuroiTV '12*, p. 45, 2012.
- [16] T. Seyed, C. Burns, M. Costa Sousa, F. Maurer, and A. Tang, "Eliciting usable gestures for multi-display environments," *Proc. 2012 ACM Int. Conf. Interact. tabletops surfaces - ITS '12*, p. 41, 2012.
- [17] "Understanding Users' Preferences for Surface Gestures."
- [18] G. Apitz and F. Guimbreti re, "CrossY," *Proc. 17th Annu. ACM Symp. User interface Softw. Technol. - UIST '04*, vol. 6, no. 2, p. 3, 2004.
- [19] D. Freeman, H. Benko, M. R. Morris, and D. Wigdor, "ShadowGuides," *Proc. ACM Int. Conf. Interact. Tabletops Surfaces - ITS '09*, p. 165, 2009.
- [20] D. Wigdor and D. Wixon, "Brave NUI World," *Brave NUI World*, vol. 36, no. 6, pp. 29–30, 2011.

- [21] J. T. K. V. K. Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindlbauer, Alexandra Ion, Shengdong Zhao, “Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for HCI,” *Lect. Notes Comput. Sci.* (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10267 LNCS, pp. 288–296, 2012.
- [22] C. Groenewald, C. Anslow, J. Islam, C. Rooney, P. Passmore, and W. Wong, “Understanding 3D Mid-Air Hand Gestures with Interactive Surfaces and Displays : A Systematic Literature Review,” *Proc. Br. HCI 2016 - Fusion*, pp. 1–13, 2016.
- [23] M. Karam, “GestureTaxonomyJuly21,” pp. 1–45, 2005.
- [24] J. Liu, “Analysis of Gender Differences in Speech and Hand Gesture Coordination for the design of Multimodal Interface Systems,” no. June, 2014.
- [25] W. A. Rogers and A. D. Fisk, “Attaining a European Reputation, 1898–1904,” pp. 645–653, 2010.
- [25] Cadoz, C., *Les réalités virtuelles*. Flammarion, Paris, 1994. ISBN: 2-08-035142-7.
- [26] McNeill, D., *Hand and Mind: What Gestures Reveal About Thought*. The University of Chicago Press, Chicago, London, 1995. ISBN: 0-226-56134-8
- [27] Kendon, A., How gestures can become like words. In F. Poyatos (Ed.), *Crosscultural perspectives in nonverbal communication*. Hogrefe, 1988, 131-141
- [28] Efron, D., *Gesture and environment*. The Hague: Mouton, 1972.
- [29] Knierim, Pascal, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. “Physical Keyboards in Virtual Reality: Analysis of Typing Performance

and Effects of Avatar Hands.” Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18 1–9.

[30] Yu, Chun, Yizheng Gu, Zhican Yang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. “Tap, Dwell or Gesture?: Exploring Head-Based Text Entry Techniques for HMDs.” Chi 2017 0–0.

[31] Yu, Chun, Yizheng Gu, Zhican Yang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. “Tap, Dwell or Gesture?: Exploring Head-Based Text Entry Techniques for HMDs.” Chi 2017 0–0.

[32] Gray, Keenan R. 2018. “Facilitating Keyboard Use While Wearing a Head-Mounted Display.” Masters Thesis.

[33] Marco, E. 2007. “UC San Diego UC San Diego Electronic Theses and Dissertations.”