<research paper>

Quantitative Comparison Between A Traditional User Interface (GUI) and A Perceptual User Interface (PUI)

</research paper>

Arjun Yadav
B.A Hons. Communication Design, *Research Methods 2021*
Indian Institute of Art & Design, New Delhi
03 December 2021.

<introduction>

In the book *Computers and Society*, Ronald M. Baecker shares an excerpt from a paper published by J.C.R Licklider [1] which states that *the hope is that, in not too many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.*

This speculative vision was discussed by many distinguished academicians over the years such as Vannevar Bush, Doug Engelbart, Ted Nelson, and Alan Kay [2]; much of which has come true today in the 21st century. Embedded computing systems have become an irreplaceable pillar of modern-day society facilitating functions necessary for communal sustenance. No longer is the traditional desktop paradigm, with a display interacted via a keyboard & mouse, appropriate in the ICT age[1]. As discussed by Lenman, Bretzner and Thuresson [3], *natural actions in human-to-human communication, such as speak and gesture, seem more appropriate for ... everyday computing, and which should support the informal and unstructured activities of everyday life.*

Primarily, interaction with a computer system requires communication through a modality or modalities which can be defined as an independent channel(s) of input/output between a human and machine. Simple modalities such as touch and the usage of GIDs[2] create a layer of insulation between man and machine that must be removed for interaction to feel more natural and efficient, bringing HCI closer to human-human interaction. As Raskin expands upon in his book, *The Humane Interface*, there is a need for interfaces to be more humane in the 21st century; that is to be *responsive to human needs and considerate of human facilities* [12].

In recent years, a new area referred to as perceptual interfaces [13] has emerged; committed to making human-computer interaction feel more natural by integrating perceptual modalities; a radical example of which can be seen in the work done by Pranav Mistry as part of the Fluid Interfaces Group at the MIT Media Lab [Fig. 1].

Fig 1: The Sixth Sense technology is a wearable, gesture-driven personal computing system with the goal of making data accessible as the "sixth sense"; Image Source: Pranav Mistry.

1. The ICT Age, here, refers to the merging of information and communications technologies in a way that has ushered the world towards a new age / type of society as discussed by Frank Walter [3].

2. A GID refers to a Graphical Input Device.; such as a keyboard & mouse.

As the world contemplates this shift from traditional interfaces to perceptual interfaces, there is a lack of any quantitative data on how much faster a perceptual interface might be when tested against a traditional interface. This paper presents a head-to-head quantitative analysis and interface design comparison between a traditional interface and perceptual interface when the content and purpose of both interfaces remain the same, but input modalities are changed.

</introduction>

<setup of the experiment>

The conducted experiment sought to compare a traditional interface with a perceptual one using a quantitative model for analysis (refer to *model to collect data*). To accommodate for the same, a simple common task and purpose, as well as a unified structure, had to be chosen.

Participants were required to fill out a form, with each question having a binary answer possibility (Yes / No). There were a total of six (6) questions adapted from the OSHA Sample Employee COVID-19 Health Screening Questionnaire [5]. These questions were chosen because of their relevance at the time and also because each question could be answered via a simple yes/no.

The test was conducted as a one-on-one test in a moderately controlled environment which was a closed classroom. The test had two parts. In the first part, participants recorded their answers using a head-tracked interface* (built using Processing, a Java environment). The interface took in input from the webcam and participants could move their head to the right or left to input answers for each question. The computer displayed output using the monitor.

Fig 2: A participant using the head-tracked interface.

*The head tracked interface is available as an open-source resource [22] to serve as the starting point for more experimentation and further research.

In the second part, participants were presented with the same set of questions and answer possibilities but had to input their answers using Google Forms, a popular interface used to take surveys. Google Forms was chosen because of its similarity in design with most other self-check-in tools that present the questions in a form-like format.

Fig 3: A participant using the Google Forms form.

For both tests, a 2019 16-in MacBook Pro was used as it was the only accessible device during the study. Impacts of the device selection are elaborated upon in the section titled, *factors to consider*.

Participants were introduced to both interfaces and provided with guidance on usage wherever needed (particularly in the head-tracked interface) before starting the test. This was done to create some kind of system image[3] for a type of interface that they had never used prior to this test, in order to match the system image that they already had of using a Google Form. Additionally, a signifier[4] (here, a paper prompt) was also added to remind participants of available affordances[5].

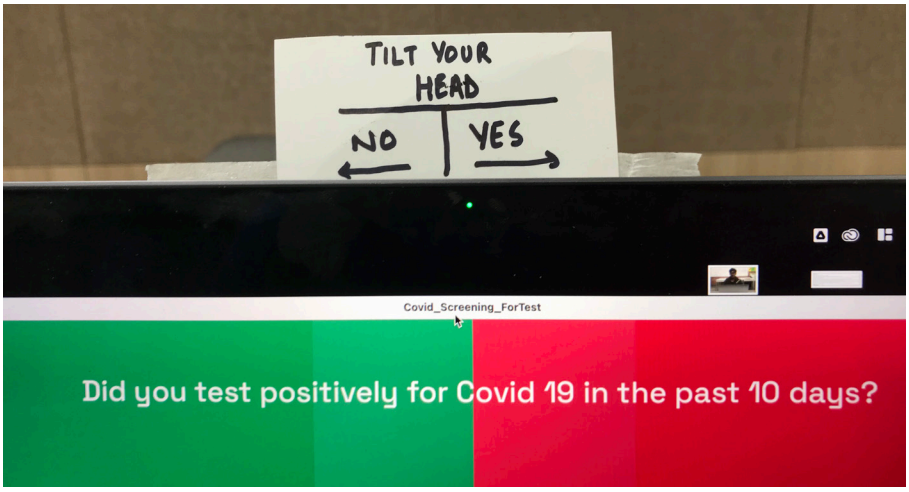

Fig 4: Signifier present during all testing sessions; enforcing the tilting of head and its subsequent affect on the interface.

3. The combined information available to a user regarding a product from all possible sources is called the system image [7].

4. Signifiers refer to perceivable signals towards possible actions and how to use them; based on Norman's definition for the same [8].

5. An affordance can be described, here, as interaction possibilities between the properties of an object and capabilities of the human [9].

Additionally, participants were instructed to read the questions and their answer choices aloud as they navigated their way through the two interfaces till completion of the form and their actions were recorded using a smartphone for the entire duration of the test [22].

</setup of the experiment>

<selecting participants>

This paper was written under a tight academic deadline of two weeks during which schools and colleges in Delhi were unexpectedly closed off [10]. Therefore, the main criteria for selecting participants became accessibility and the minimum number of participants required to validate a study. Therefore, a group of five students at IIAD contributed to this study. The number five was chosen based on a study published by Jakob Nielsen for the Nielsen-Norman Group [11].

Participants were of the age group 20-28 (ages being 20, 20, 21, 21, 28) years with fairly similar levels of digital literacy as they belonged to the same academic environment and discipline of study.

In further testing under more favourable conditions, an essential factor for screening participants would be digital literacy as it was observed in the test that this particular group had a high amount of digital literacy which definitely impacted the ease with which they interacted with the computer.

</selecting participants>

<model to collect data>

In *The Humane Interface*, Raskin elaborates upon models used for the quantitative analysis of interfaces [14]. A commonly used model known as the GOMS Keystroke-Level Model, introduced by Card, Moran and Newell in 1983 [15], was adapted for this experiment. The inventors of the KLM model state that the total time it takes to complete a task on a computer system is the sum of elementary gestures that the task comprises [19]. While other models prove to be more accurate and precise, the GOMS Keystroke-Level model was the simplest model to execute and easier to adapt for the analysis of a perceptual interface as well.

As the Keystroke-Level model was originally developed to analyse interface efficiency of a traditional Graphical User Interface, interacted via a keyboard and mouse, certain adjustments to the model were made to accommodate the nature of the perceptual interface built for this experiment. The model used certain mnemonics (based on the GOMS-KLM mnemonics) to track different stages of interaction, each of which has been broken down in the section titled, *mnemonics used and what they stand for.*

</model to collect data>

<design of the interfaces>

The design of the two interfaces has a considerable impact on the data. Therefore, in many ways, the study cannot be limited to quantitative metrics of interaction without considering the impact of the interface design.

The two interfaces differ majorly from a design viewpoint because of the difference in input modalities, even though the contents of the interface are the same. There exists a mouse pointer on both interfaces (a bounding box tracking your head in case of the head-tracked interface), the question (*n*) and the two possible answer choices (can be considered the only two buttons on each interface); as seen below in Fig 5&6.

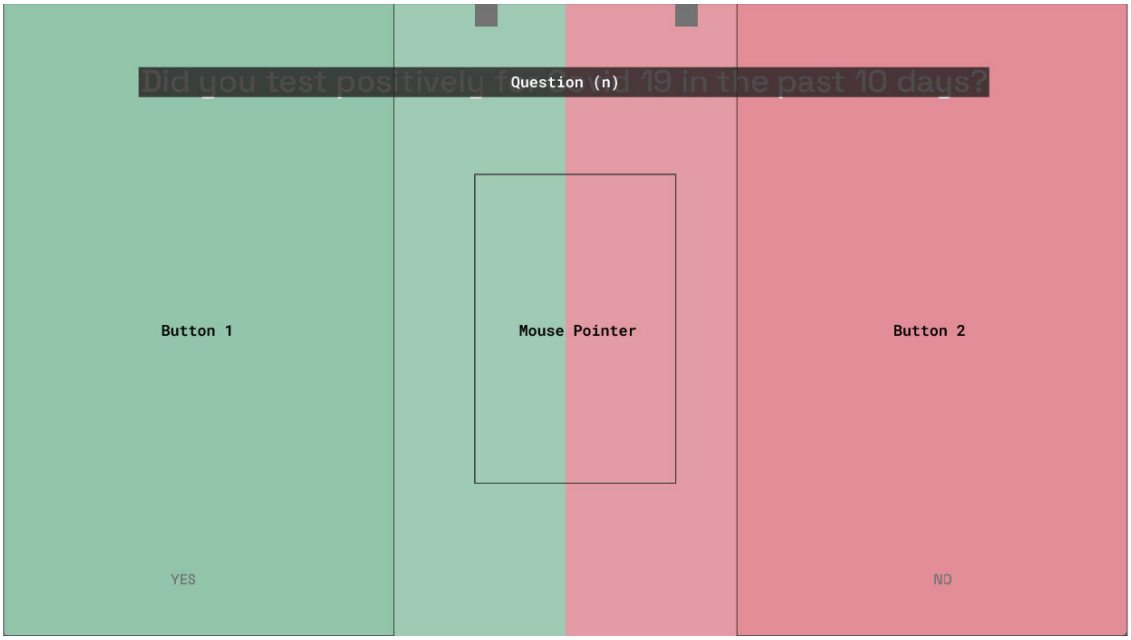

Fig 5: The head tracked interface built on Processing [7]; a thing to remember is that the mouse pointer is controlled with the head position.

Fig 6: The Google Forms interface part of the free Google Docs Editor suite by Google.

A major argument to be considered when studying the design and its impact on the usability data is the implication of Hick's Law impacting the times it took to navigate through different sections of either interface. Hick's law states that *the more stimuli to choose from, the longer it takes the user to make a decision on which one to interact with* [16].

The head-tracked interface works in such a way that questions come individually, one after the other, ensuring that users only have the necessary information on screen to comprehend. This means that the structure of the program works in such a way that the display of the next question is dependent on the user returning to the neutral state post-answering (entering input).

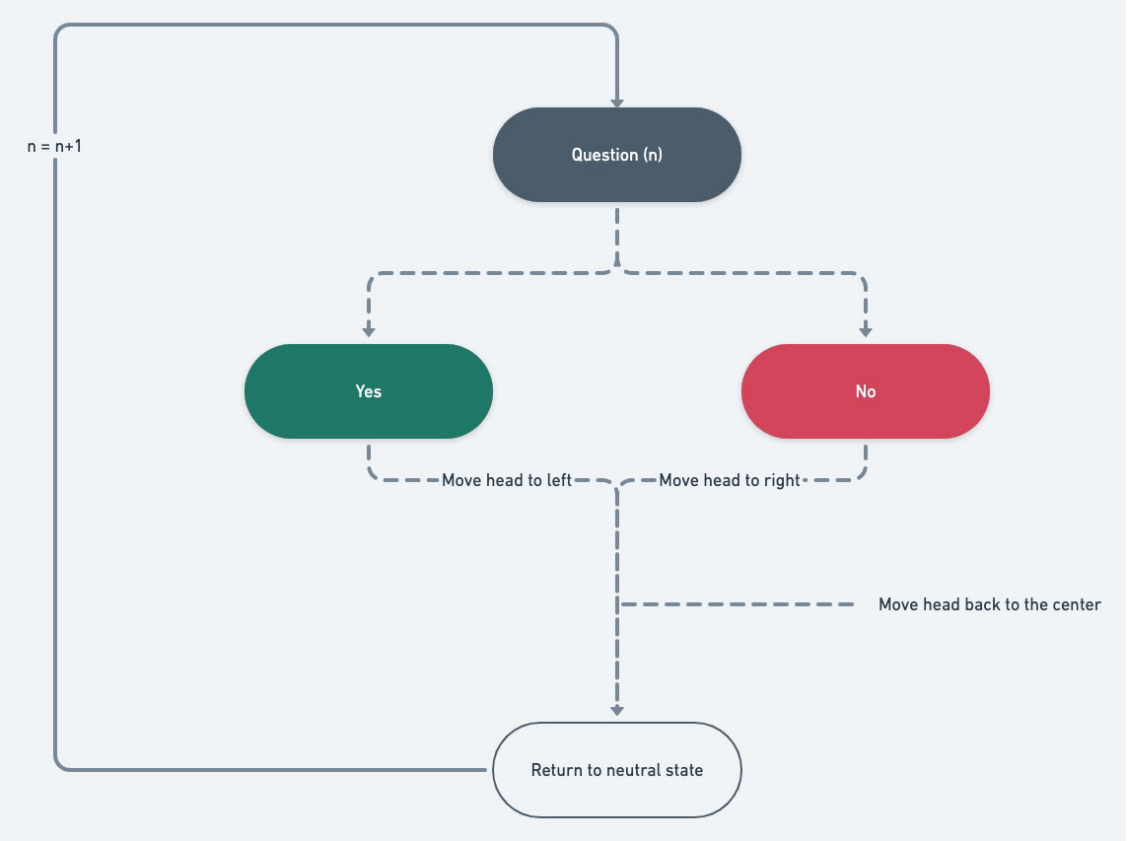

Fig 7: Algorithm flow of the head-tracked interface.

On a Google Forms form, however, a list of questions can be seen, increasing the amount of information that a user has to comprehend in a single glance.

However, it should be noted that each question and answer possibilities (buttons) are separated into individual cards. Therefore, user attention is retained on the active card (the question the user is on), which is precisely how the head-tracked interface works as well; the comparision of which can be seen in Fig 8.

Therefore, it cannot be concluded that Hick's Law corrupts the data in a major way as the number of elements that grab a user's attention at one point in time remain the same.

What remains now is the presentation of the elements on the interface. This is the major point of differentiation that can account for some impact on the data. However, this is a necessary point of differentiation as the interface had to be designed keeping the modalities of interaction between the machine and man in mind. A Google Forms form with the same design will be extremely difficult to navigate using a head-tracked interface and vice-versa. Therefore, for the head-tracked interface, the design was kept to the bare minimum; making sure that only the necessary elements are included in the interface steering clear of additional colours and other such elements that could impact usability.

Fig 8: Cropped screenshots of both interfaces; top half being the head-tracked interface and the bottom being the Google Forms form in this image.

</design of the interfaces>

<mnemonics used and what they stand for>

Head Interface

- S refers to the time it took a participant to see the question before starting comprehension; time between the display of the question on the screen and the starting of comprehension (C).

- C refers to the time it took a participant to comprehend the displayed question; time marked by the participants' audible statement of reading the question aloud from start to end.

- P refers to the time it took a participant to move their head to the preferred answer zone; duration of movement of the head from stationary till the point the computer recognized input*.

- M refers to the time it took a participant to mentally prepare for the next step (could include thinking about the answer); time between the movement of their head from the answer zone to the neutral zone.

- R refers to the time it took the computer to respond to input; time it took the filled screen (red/green) to return to neutral and display the next question.

Since M & R had overlapping times, they were clubbed together as M+R while recording the data.

Form

- C refers to the time it took a participant to comprehend the displayed question; time marked by the participants' audible statement of reading the question aloud from start to end.

- P refers to the time it took a participant to point to a position on the

display; time taken to move the mouse pointer from one position to the desired position when the intended action is to point to an object on the interface.

- H refers to the time it took a participant to move their finger(s) from the air to the trackpad; time taken to touch the trackpad from the air.

- K refers to the input gesture of a tap; time taken to input a tap (marked by the release-homing-release of the finger, expanded in Gestures Sec 4.2).

- M & R, in this interface, were impossible to calculate as it was hard to track the definite starting point for M and there was virtually no time recorded for R as all questions were on a single page.

Interaction in the GOMS model results in equations of interactions that are further elaborated upon in the next section, *interaction with the interfaces.*

</mnemonics used and what they stand for>

<interaction with the interfaces>

The interactions were different for both interfaces because of the difference in input modality as well as the either program processed input and displayed output. However, there are certain common patterns of interaction and movement that lead to similarities in the conceptual model[6] required to navigate either interface.

The head-tracked interface used a light bobbing gesture which involved the movement of the head from left to right; when rotated from the axis of the neck. A center to left/right movement resulted in the input of a desired answer (yes/no) and a movement from left/right to the center resulted in the computer reaching a neutral stage and displaying the next question.

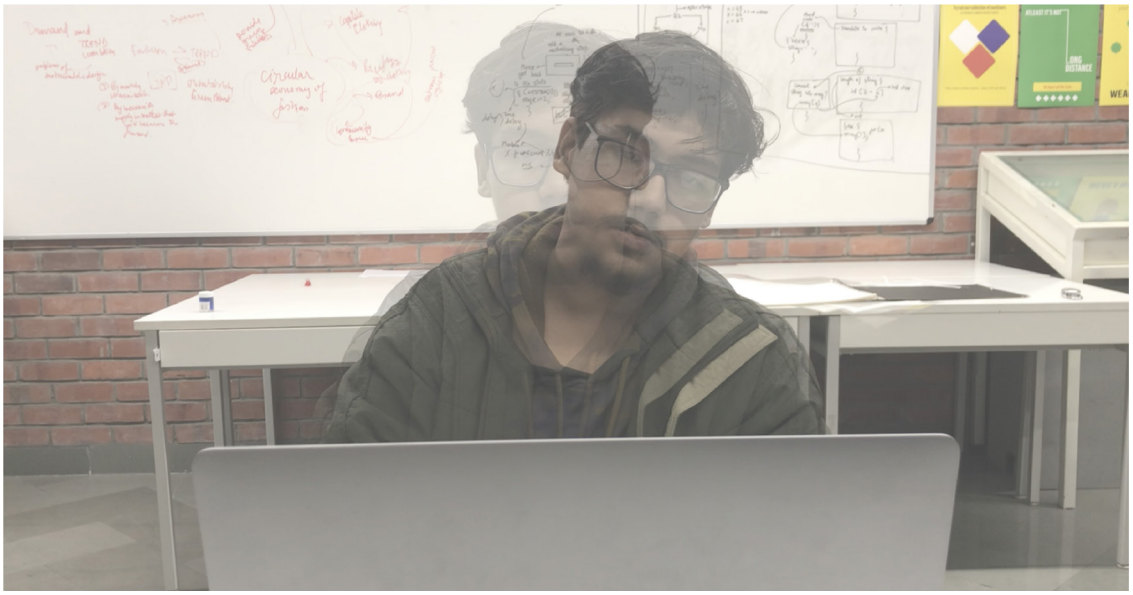

Fig 9: A participant's head movement in order to input "yes" to the head-tracked interface.

Interaction patterns for the traditional interface were a little more complex. There were certain set patterns of interaction that are common in traditional personal computing on a laptop.

Each individual step has been broken down earlier in the explanation of mnemonics [refer to *mnemonics and what they stand for*]. However, a

6. Here, I refer to conceptual model as a mental model that people form when interacting with an interface that determines their interaction pattern with the same [17].

combination of steps can be referred to as gestures equivalent to those of the head-tracked interface, which is what I shall expand upon below.

A single tap, post Pointing (using a lateral/vertical/diagonal movement) to the desired button on the screen was used to feed input into the computer. The end of this gesture was marked by the release of the finger from the trackpad, as a tap without release would be called a mouse press which would not provide the same result as a tap. This can be broken down into the following equation:

*Tap = P + K**
*where K is actually (Release + Homing + Release).

After input, as participants moved on to the next question, it was observed that they used the mouse pointer as a tool to direct their attention (further discussed in the section titled *findings*). This can be broken down into the following equation:

Pointing to next question = H + P

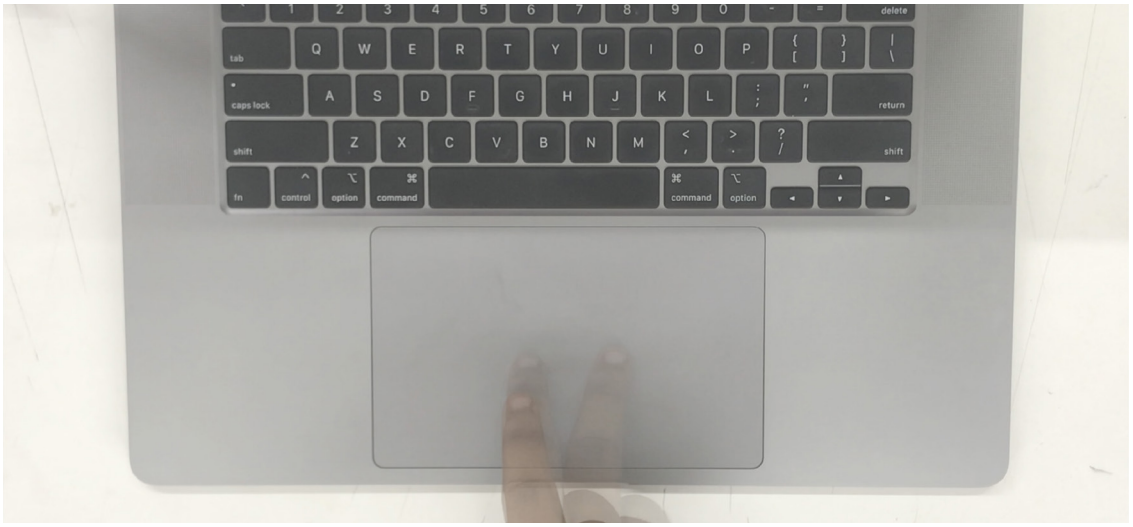

Fig 10: A sample range of movement considered for the pointing stage; lateral movement of the finger.

There was an additional gesture in the traditional interface known as scrolling. This meant moving data on-screen upwards through a downward scroll of the trackpad to reveal the remaining questions. This can be broken down into the following equation:

Scrolling = H (two fingers) + P (both fingers)

</interaction with the interfaces>

<factors to consider>

- The most glaring limitation is the selection of participants. In the given time frame and conditions [10], I was only able to test with 5 users who had fairly similar digital literacy levels in my immediate proximity; which is the IIAD College Campus in Okhla, New Delhi.

- The tests were conducted on a 2019 16-in MacBook Pro using the inbuilt trackpad (with 5 units of Tracking Speed on the macOS) and the 720p FaceTime webcam with a display refresh rate of 60fps. The input devices used also have an impact on the input times for both, the head-tracked interface and the Google Forms form. Personal computing devices have

the liberty of having a customised DPI which could increase/decrease the speed of pointing to a desired location for individuals familiar with the computer system. However, interfaces that exist in a social context; with a single computer being used by multiple people (such as in this test) would not have this same liberty.

</factors to consider>

<findings>

It was observed that the time taken for participants to feed input into the computer post comprehension (C) for one individual question (the average of times taken for Q1 and Q3) was faster in the head-tracked interface by 0.54 seconds, which is an improvement in input time by 54%.

There were also an additional 1.625 steps that a participant had to perform when moving from one question to the next post answering the question (Input). These included possible movements where participants used the mouse pointer as a tool to direct their attention as well as scroll to move onto the next question.

Another interesting observation was that even though the device used for testing displayed 4 questions in the first glance and the list of questions could be navigated through a single scroll, participants subconsciously scrolled after almost every question, wanting to keep each question card at a fixed point of attention.

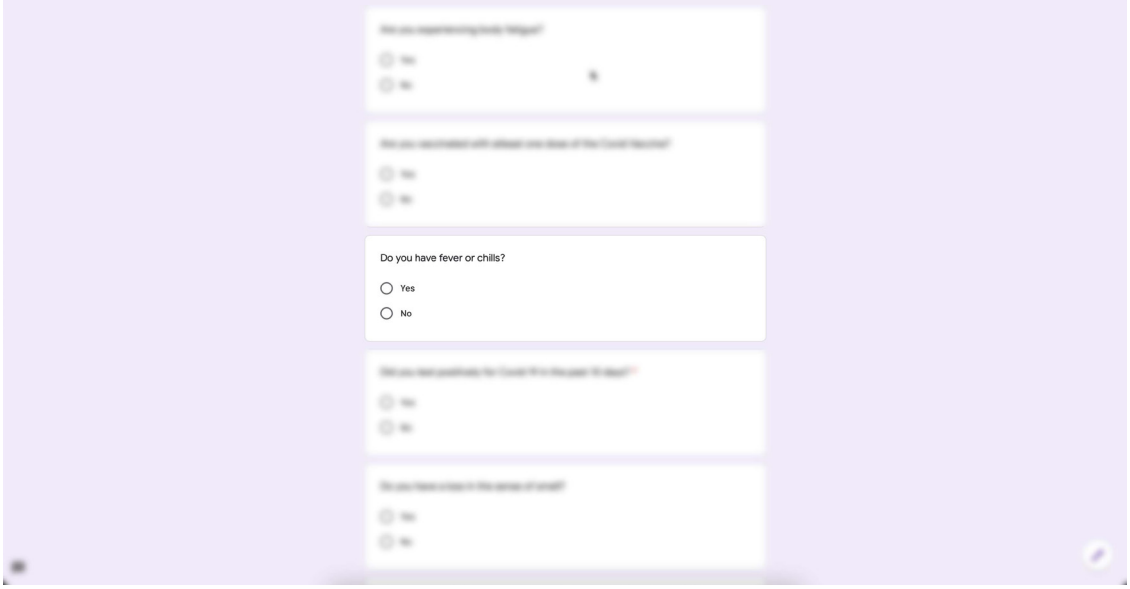

Fig 11: A hypothesized visualisation of the "fixed point of attention" as mentioned above. This assumption can be validated via a device that can facilitate gaze-tracking.

The final piece of analysis to discuss is that there was a 37% improvement in response time for Q3 when compared to that of Q1 for the head tracked interface and a 35% improvement in the traditional interface for the same. This is an interesting piece of data as before participating in this study, none of the participants had any prior experience of using a computer-vision based interface. One could logically hypothesize that the learning curve of using a computer vision based interface was almost negligible in this particular scenario and, surprisingly, even lower than that of a platform which participants used frequently (Google Forms). This could further mean that the gesture of moving one's head in the direction of

yes/no replicated a fairly common gesture of nodding as a response in human-human communication; meaning that the head-tracked interface inadvertently used a natural gesture to navigate through.

Using followup questions, it was discovered that 100% of the participants found the head-tracked interface easier to use. However, in two cases, it was found that using the traditional interface felt more natural to them as they were used to it / accustomed to using a trackpad. In order to truly understand which is the more natural or humane interface in this case, one must conduct a similar experiment with people who have low levels of digital literacy, as the barrier of a hardened system image would not be a factor when comparing the ease of use of both interfaces.

</findings>

<conclusion>

The decrease in input time for a perceptual interface implies that, in this particular scenario, a PUI can lead to a faster completion time for the task at hand. This means that a change in input modality could be particularly helpful when time to complete a task is essential. However, it must also be explicitly stated that there was a 0% error rate throughout my test only because it was conducted in a controlled environment. Using a computer-vision based interface in a social setting can undeniably lead to issues with facial recognition and other such problems which majorly have to deal with the technology itself and the algorithms used. This factor was not considered in this study.

Furthermore, the presence of a signifier to guide head movements is something that an interface designer may or may not want to include in their interface. Whether the interface naturally affords to be used in the intended way still remains unclear as it was not an objective of this particular research project.

The total effort it took to complete a task is less in the case of a perceptual interface because of the difference in total step count necessary to complete a task. This is a major difference between traditional and perceptual interfaces as it changes the pattern of human-computer interaction and justifies why perceptual interfaces feel more natural and, therefore, also become more efficient and faster to use. Using gestures that mimic human-human communication can also aid in this effort, such as the gesture to nod in the head-tracked interface.

Finally, it is imperative to understand that the modality defines the interface design and its subsequent user experience in a massive way. Existing GUI interfaces can unquestionably be enhanced in their design when the primary modality of interaction is changed. With a combination of natural (perceptual) modalities and an interface that embraces these modalities, human-computer interaction can strive to become more humane; and serve their societal functions more efficiently.

This paper, validated by the conducted experiment, can successfully conclude that simple interfaces, where there exist a lesser number of variables, can improve usability, functionality and experience with the integration of a perceptual modality by a significant margin.

</conclusion>

<bibliography>

[1] Ronald M. Baecker, Computers and Society,: Modern Perspectives, Oxford University Press, 2019, p. 22

[2] Ronald M. Baecker, Computers and Society,: Modern Perspectives, Oxford University Press, 2019, p. 42

[3] Frank Walter, Information Society Theories: Third Edition, Routledge, 2019, pp. 9-11

[5] Occupational Safety and Health Administration, Sample Employee COVID-19 Health Screening Questionnaire, www.osha.gov, 2019

[7] Don Norman, The Design of Everyday Things: Revised and Expanded Edition, Basic Books, 2013, pp. 31-34.

[8] Don Norman, The Design of Everyday Things: Revised and Expanded Edition, Basic Books, 2013, pp. 13-19.

[9] Don Norman, The Design of Everyday Things: Revised and Expanded Edition, Basic Books, 2013, pp. 10-13

[10] BBC News, Delhi smog: Schools and colleges shut as pollution worsens, 2021, online: https://www.bbc.com/news/world-asia-india-59258910

[11] Jakob Nielsen, Why You Only Need to Test with 5 Users, 2000, Nielsen and Norman Group, https://www.bbc.com/news/world-asia-india-59258910

[12] Jef Raskin, The Humane Interface, Addison - Wesley, 2000, p. 6

[13] Matthew Turk and Mathias Kölsch, Perceptual Interfaces, University of California Santa Barbara, 2004, p. 455, https://sites.cs.ucsb.edu/~mturk/pubs/TurkKolsch2004.pdf

[14] Jef Raskin, The Humane Interface, Addison - Wesley, 2000, pp. 71-78

[15] Jef Raskin, The Humane Interface, Addison - Wesley, 2000, p. 72

[16] Mads Soegaard, Hick's Law: Making the choice easier for users, Interaction Design Foundation, 2020, https://www.interaction-design.org/literature/article/hick-s-law-making-the-choice-easier-for-users

[17] Don Norman, The Design of Everyday Things: Revised and Expanded Edition, Basic Books, 2013, pp. 25-31.

[18] Suman Bhandary, Letterform Theories: Gerrit Noordzij's moving pen, University of Reading, 2019.

[19] Jazer Kenaz Chand, Bi Poetry, Pearl Academy Delhi, 2020.

[20] Jef Raskin, The Humane Interface, Addison - Wesley, 2000, p. 73.

[21] Neri Oxman, Age of Entanglement, Journal of Design and Science (JoDS by the MIT Media Lab), 2015.

[22] The data sets generated for the test along with the source-code of the interface: https://drive.google.com/drive/folders/1Yd9rm1yLGwffcS3UsO4Mi15S6NGW6BKp?usp=sharing

Apart from the sources listed above, there exist people without whom this paper would have never reached an end. I'd like to acknowledge the efforts of my mentors, Ms. Prachi Mittal and Mr. Suman Bhandary as well as my peers: Nikhil Shankar, Pratishtha Purwar, Alina Khatri, Atreyo Roy, Kriti Agarwal, Navya Baranwal, Harshvardhan Srivastava and Natesh Subhedar who helped me shape the contents of my argument and allowed me to test my interfaces.

I'd also like to acknowledge the library at IIAD as well as Mr. Paramjit Singh and Mr. Natesh Subhedar for their help in my research.

The last bit of gratitude goes out to the efforts of Daniel Shiffman, the Processing team and Peter Abeles for their respective contributions to open-source software and the propagation of related knowledge without which this entire paper would have remained a hypothesis.