

RESEARCH ARTICLE

Enhancing the usability of low-cost eye trackers for rehabilitation applications

Rahul Dasharath Gavas, Sangheeta Roy, Debatri Chatterjee*, Soumya Ranjan Tripathy[‡], Kingshuk Chakravarty, Aniruddha Sinha

Embedded Systems & Robotics, TCS Research and Innovation, Tata Consultancy Services, Kolkata, India

[‡] Current address: Department of Signal Processing, Tampere University of Technology (TUT), Tampere, Finland

* debatri.chatterjee@tcs.com



OPEN ACCESS

Citation: Gavas RD, Roy S, Chatterjee D, Tripathy SR, Chakravarty K, Sinha A (2018) Enhancing the usability of low-cost eye trackers for rehabilitation applications. PLoS ONE 13(6): e0196348. <https://doi.org/10.1371/journal.pone.0196348>

Editor: Enkelejda Kasneci, University of Tübingen, GERMANY

Received: September 12, 2017

Accepted: April 11, 2018

Published: June 1, 2018

Copyright: © 2018 Gavas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript.

Funding: Tata Consultancy Services Ltd. (TCS) provided support in the form of salaries for all the authors and provided infrastructural support and subjects for data collection, decision to publish and preparation of the manuscript. The clearance on ethical issues for handling and analysis of the data collected has been acquired from Institutional Review Board of TCS. The specific roles of these authors are articulated in the 'author contributions' section.

Abstract

Eye tracking is one of the most widely used technique for assessment, screening and human-machine interaction related applications. There are certain issues which limit the usage of eye trackers in practical scenarios, viz., i) need to perform multiple calibrations and ii) presence of inherent noise in the recorded data. To address these issues, we have proposed a protocol for one-time calibration against the “regular” or the “multiple” calibration phases. It is seen that though it is always desirable to perform multiple calibration, the one-time calibration also produces comparable results and might be better for individuals who are not able to perform multiple calibrations. In that case, “One-time calibration” can also be done by a participant and the calibration results are used for the rest of the participants, provided the chin rest and the eye tracker positions are unaltered. The second major issue is the presence of the inherent noise in the raw gaze data, leading to systematic and variable errors. We have proposed a signal processing chain to remove these two types of errors. Two different psychological stimuli-based tasks, namely, recall-recognition test and number gazing task are used as a case study for the same. It is seen that the proposed approach gives satisfactory results even with one-time calibration. The study is also extended to test the effect of long duration task on the performance of the proposed algorithm and the results confirm that the proposed methods work well in such scenarios too.

1 Introduction

In recent years, eye tracking is gaining huge importance for diagnosis and screening [1] of various medical conditions, home-based rehabilitation [2] and human-computer applications [3] due to its unobtrusive nature. Eye tracking is also an important method for analyzing different cognitive functions [4] associated with variety of tasks like reading, writing, visual searching, driving and so on. Non-invasive eye trackers can also be used to study infant cognition [5] in unconstrained, naturalistic environment. However, the accuracy or the robustness of such applications mostly relies on the quality of the data collected. Noisy eye movement data leads to misleading interpretations and outcomes.

Competing interests: Tata Consultancy Services Ltd. (TCS) provided support in the form of salaries for all the authors and provided infrastructural support and subjects for data collection, decision to publish and preparation of the manuscript. The clearance on ethical issues for handling and analysis of the data collected has been acquired from Institutional Review Board of Tata Consultancy Services Ltd. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Statistics shows that about 8% of world's total population is the aged population [6], most of which are suffering from some ailments leading to cognitive decline [7] affecting their oculomotor response. Lagun et al [8] showed that Visual Paired Comparison (VPC) task usually provides insights to memory impairments associated with mild cognitive impairment which often progresses to Alzheimer's disease. Saccadic eye movements can also be used to quantify motor impairments in Parkinson's disease (PD) [9]. Regular monitoring of eye movement definitely plays a crucial role in assessing cognitive states of such patients. On the other hand, in order to be an ideal choice for home-based rehabilitation applications, the eye tracking device should be portable, easy to use and most importantly affordable. However, such low-cost devices are majorly low in resolution, thereby compromising on the quality of the data [10] recorded. This is usually handled to some extent through a calibration phase that needs to be performed at the beginning of each session and most of the times, in between the experiments too [11]. This is termed as multiple calibration, which is a cumbersome and repetitive process. Moreover, achieving good calibration is a major challenge for patients (with Stroke, Parkinson's disease, Dementia, Schizophrenia etc.) and infants [10] as they lack the patience and capability to gaze on a fixed point for longer duration. The process also leads to exhaustion or loss of engagement during the actual task that is performed after the calibration phase. Hence, there is a need for establishing a one-time calibration protocol for experiments/tasks targeted for the above discussed participant groups. Another major problem associated with these low-cost eye trackers are the presence of huge inherent noise in the recorded data. Even if we somehow manage to achieve a good calibration score, the quality of data acquired after such sessions is highly susceptible to inherent noise which is a result of head movements, glitches in the eye tracker sensor algorithms, lightning conditions, and so on. In addition to this, the subject-specific variances due to drift, micro-saccades, tremors, etc. are also present in the collected eye gaze data. These issues actually gives rise to certain noise in the collected eye tracker data which can be broadly classified into 2 major classes – variable and systematic errors [12, 13] as shown in Fig 1. The former refers to the dispersion of the gaze coordinates for a given target (Fig 1(a)) and the latter refers to the drift from the target location (Fig 1(b)). These errors are present irrespective of single or multiple calibrations; but the degree of systematic error is larger in the former in comparison to the latter [14]. Variable error is indicative of the lack of precision and the systematic error is indicative of the lack of accuracy [12]. The non-systematic/variable errors are mainly related to physiological characteristics of a participants' eye and fatigue; and head motion [15]. Changes in screen illumination, participants' ethnicity and operating distance from the screen [16], etc. also add to the degradation of the eye tracker accuracy. High-end eye tracking systems like Tobii also suffers from such errors. Efforts to denoise Tobii are reported in [14]. However, such high-end eye trackers are not mass-deployable for home-based rehabilitation owing to their high cost; whereas, in low resolution eye trackers, the amount of system generated noise is greater than the natural (inherent) noise. Without the removal of these errors, it is difficult to use eye trackers for practical applications, especially in human-computer interaction (HCI)-based clinical applications [17–19].

Thus, it can be concluded that the two major challenges that reduce the usability of low-cost eye trackers in various medical or rehabilitation applications are as follows: 1) need for multiple calibration and 2) inherent variable and systematic errors. The present study proposes an algorithm which takes care of the noises associated with eye trackers and also a protocol in order to avoid the overhead of multiple calibration is devised.

The present work is the detailed algorithmic version of the simple eye tracking noise cleaning approaches that we proposed earlier [20] with a larger population set, additional metrics to study the performance of the proposed algorithms on variable and systematic error

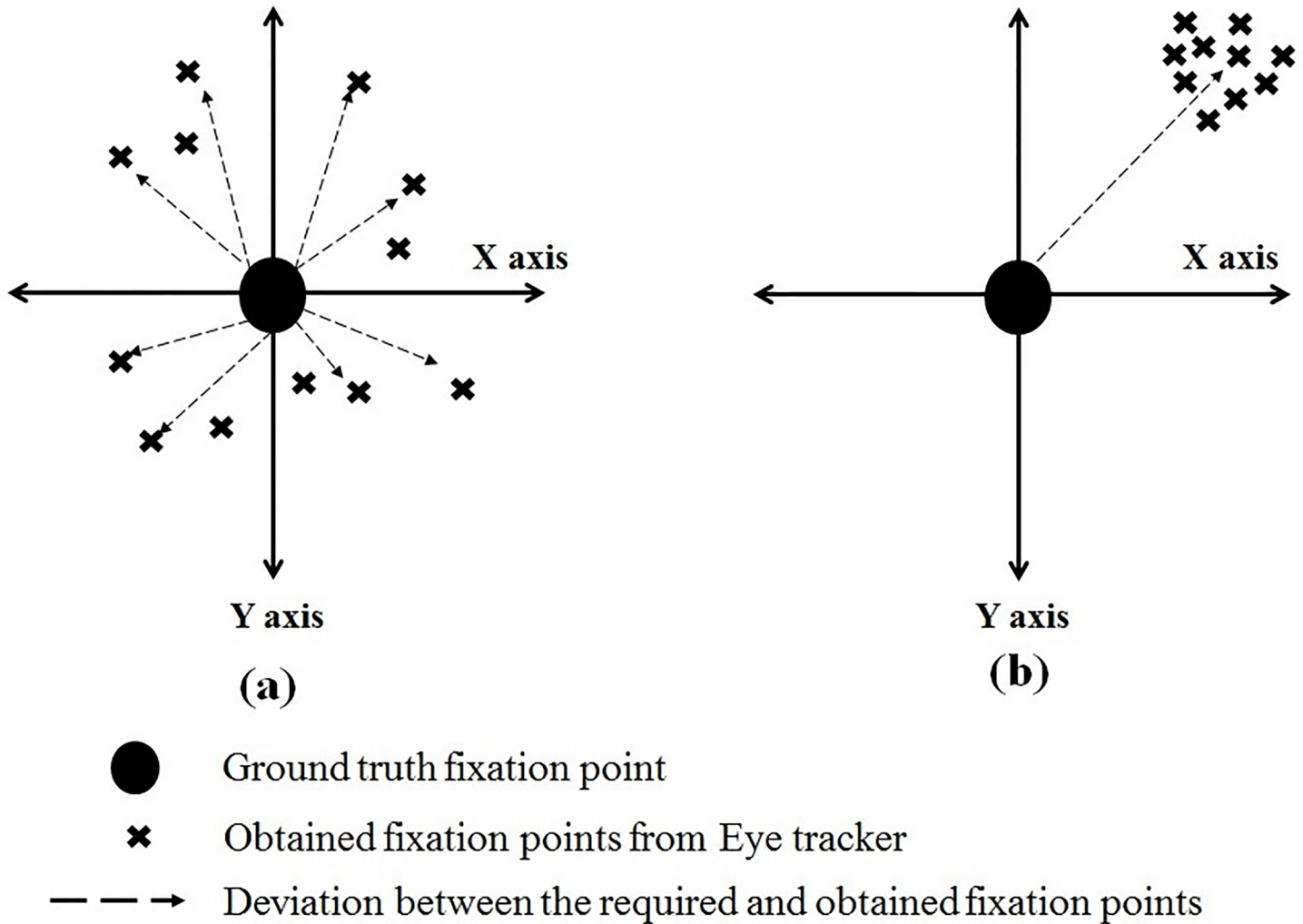


Fig 1. Two types of errors associated with eye trackers, (a) Variable and (b) Systematic error.

<https://doi.org/10.1371/journal.pone.0196348.g001>

correction, detailed analysis of the supervised and unsupervised approaches and the study of the effect of longer duration tasks on the proposed approaches.

2 Related work

Eye trackers are mainly categorized into 2 types based on their design features, namely remote (nearables [21]) and wearables (head mounted [22]). Each one of these comes with its own set of advantages and disadvantages. For instance, the nearable ones are unobtrusive in nature but are less efficient in comparison to the wearables. Also, the participants become more cautious about the sensors that they are wearing during the actual experiment. One of the most robust wearable eye trackers uses contact lens, wherein the tracking mechanism is embedded into the lens [23]. However, it is costlier in comparison to other eye tracking devices and also its complexity makes it less user friendly. Electro-oculogram (EOG)-based eye tracking is yet another popular means of detecting eye movements by acquiring the minute changes generated by the

corneal-retinal potentials of the eyes [24]. This method is highly vulnerable to various electrical noises and drift errors. Also, the complex circuitry consisting of wires and gel-based electrodes make it less appealing for practical scenarios. For large-scale deployment, the device should be low-cost and unobtrusive in nature. Video-based and infrared-based eye trackers are popularly used as unobtrusive means of eye tracking. Video-based eye tracking is a popular technique [25] in which a camera focuses on the eyes and the eye ball movements are recorded. One variant of this as proposed by Zhang et al. [26] is basically an appearance-based method which does not require calibration, but the accuracy of detection is highly dependent on ambient lighting conditions; thereby degrading its performance in real time scenarios. Basilio et al [27] proposed a calibration free method but it faces challenges in real life applications due to head and body movements. The accuracy of the method is less due to the following reasons: 1) severe lighting conditions owing to the head movements towards bright areas of the environment; 2) distortion due to wireless transmission of the video data and 3) absence of user calibration. Therefore, the accuracy of video-based eye trackers is compromised due to head movements, which are prevalent during longer duration experiments, thereby limiting the usage in short-term experimental sessions only. The infrared-based methods are less complex, cost-effective [28] and un-obtrusive in nature [29]. Subject-specific attributes like eyelashes covering the pupil, eye glasses or contact lenses, physiological characteristics of the eye like additional dark spots on the iris, interferes with the pupil detection algorithms [30]. In addition to these, the factors like changing illumination, recording errors, motion blur, rapidly changing illumination due to the fast movement of the participant (for instance, while driving) also adds to the errors in pupil position measurement [30]. There have been attempts to use low-cost infrared eye trackers in HCI-based applications [3]. These low-cost devices are basically lower in resolution and calibration phase plays a major role in determining the quality of the data. Multiple calibrations seem an attractive means but accomplishing it, is often cumbersome and exhaustive process [10]. Bereft of the modes of the calibration, the inherent noise namely, variable and systematic, poses major challenge for using eye tracker data.

In general, filtering-based approach is used for excavating the variable error. Most of the filtering-based approaches remove the abrupt fluctuations in the gaze data, thereby smoothing the overall signal. Some approaches design low pass filters, as suggested by Olsson et al. [31], which use both offline and online filtering to remove the noise. In the offline approach, the fixation data is extracted from the raw noisy data using sliding window approach, whereas for online approach, it estimates the filtered data by considering the mean of previously estimated position data. Many researchers proposed various techniques to compute the window size [32, 33] to estimate the fixation data. An advantage of this approach is that they allow increasing the window size depending on the application. In [34], the authors proposed a hybrid filter. It is composed of several linear FIR (Finite Impulse Response) sub-filters and finally, it performs a median filter operation over the outputs of sub-filters. The advantage of this approach lies in the ability to preserve the sharp fluctuations by attenuating the noise to some extent, whereas the fluctuations in the signal are suppressed considerably in linear low-pass filtering methods. The accuracy of these methods depends on several parameters, such as window length, fixation detection threshold, which in general are difficult to estimate as these parameters largely depend on the magnitude of the noise present at any instant of time. Some works suggest using the Kalman filter for denoising the data. It eliminates the need for storing previously observed data at each step of the filtering process. Sauter et al. [35] have proposed eye-movement detection using Kalman filter. Many authors [36] [37] have extended the Kalman filter for identifying different eye movements based on their applications. For example, in [37], the authors used Kalman filter to classify different types of eye movements and to reduce sensor lag through eye movement prediction. On similar grounds, Komogortsev et al. [38], used

an attention-based Kalman filter, which aims at reducing the noise in addition to minimizing the delay between eye gaze-based systems and displayed data for designing an interaction model based on eye movement language token.

One of the popular methods of handling the systematic error is based on the concept of extraction of 'required fixation location'. If the region on screen where the participant is gazing is known, then correcting the discrepancy in the gaze data and the ground truth is handled by estimating the amount and the direction of the drift [12]. The major shortcoming of this approach, however, lies in the fact that the error signature need not be constant throughout the experiment [39]. The error varies with sessions due to head movements, fatigue of the participant, screen illumination, changes in the distance from the screen and the ethnicity of the participants [16], etc. Also, the concept of 'required fixation location' does not apply in most of the real life scenarios, i.e. it is not always possible to know the ground truth of the gaze data. Another major approach is based on the principle of 'closest stimulus' [14] which applies the principle of annealed mean shift algorithm. This method suffers severely when the fixation has many stimulus points around it and the closest point might not be the desired target location. Also, the fact that calibration error is sometimes location dependent on the screen, the performance of this method is thus compromised [39]. Vadillo et al. [39] have proposed a linear transformation (LT) algorithm to correct the systematic error by using the concept of 'probable fixation location', which is more likely to be prevalent in practical cases. However, this method too does not retain the nature of the drift, rather it deals with the discrepancy between the target and the gaze data.

In case of human eye movements, there are inherent noises due to drift, micro-saccades, tremors, etc. However, along with these artifacts, there are noises like the variable and the systematic errors in the eye tracker data. In low resolution eye trackers, obtaining clear demarcation between the device imposed and naturally generated error, is difficult. Moreover, our aim in the current use-case is to handle fixations, and not to detect micro-saccades. The naturally generated noise is constant bereft of the quality of the eye tracker. But, in low resolution eye trackers, the characteristics of device generated noise and the natural noise closely resembles each other. Without the removal of these errors, it is difficult to use the fixation information in controlling HCI-based applications.

3 Design of stimulus

In the present study we have designed i) a set of stimulus to be used for calibration and ii) two test stimulus for evaluating the accuracy of the proposed methodology with respect to noise removal.

3.1 Calibration stimulus

Calibration is an important phase in eye tracking to collect data in order to map the coordinates of the pupil movements in the eye-video to that of the coordinates in the stimulus space. Hence, the challenge is to gather data from as many known locations as possible but with least mental effort on the participant and ensuring sustained attention on the target [40]. Blignaut [40] proposed a calibration scheme which involves collecting the data during smooth movements of the eye, termed as smooth pursuit. Thiago et al [41] proposed a dynamic calibration scheme called *CalibMe* which uses eye movements for collecting data during calibration. This method allows free head movements during calibration. In this work, we have used a simple calibration phase, which has static calibration points appearing at pre-defined positions and then it moves smoothly by generating smooth pursuits. However, we haven't used the smooth

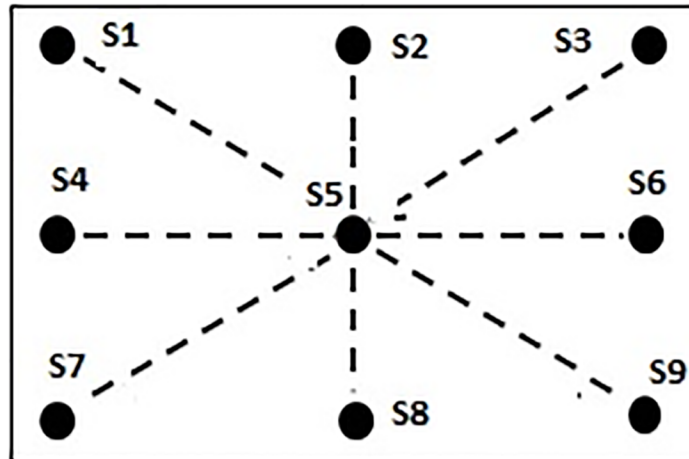


Fig 2. Schema of the designed calibration stimulus with the calibration point having diameter 0.657° and moving at a constant speed of $1.92^\circ/\text{sec}$ on the screen.

<https://doi.org/10.1371/journal.pone.0196348.g002>

pursuits for the calibration and the readers are free to use any of the above discussed calibration schemes depending upon their applications and target participants.

For the calibration purpose, 4 sets of stimuli are designed and developed using Pygame [42] (Fig 2). The stimulus consists of a tiny ball, having a field of view (FOV) of approximately, 0.657° , moving at a constant speed of $1.92^\circ/\text{sec}$ on the screen. The FOV of 0.657° corresponds to the tiny ball having a diameter of 20 pixels viewed at a distance of 60 cm. In order to cover the entire screen during the calibration phase, the ball moves in horizontal, vertical and in 2 diagonal directions as shown in Fig 2. During each of these movements, the ball stops for 5 seconds at nine positions, shown as dark spots (S1 through S9) in the Fig 2, and then it moves again along the path shown as dotted lines. The size of the ball is deliberately kept small so that the participants can easily fixate at the center of the ball. The participants are supposed to gaze carefully at the ball while their gaze data are logged using the eye tracker. In total, we have 9 static points, S1 through S9 whose corresponding fixations are used for noise correction.

3.2 Test stimulus

The test stimulus is derived from standard psychological tests and redesigned in order to test the robustness of the proposed algorithms. The test stimulus needs to be designed in such a way that it covers a broader spectrum of psychological test batteries. In this work, two stimuli are designed to test the accuracy of noise correction—1) recall-recognition (RR) [43] and 2) number gazing task (NG) (analogous to the Digit-Symbol Substitution Test (DSST) [44]) using Pygame package. A summary of the stimulus is provided in Table 1. It is to be noted that the systematic and variable errors are independent of the stimulus. The NG task contains the stimulus points (numbers through 1-9) only in a single row, whereas, the RR task presents the stimuli (words) in multiple rows. Variation in the inter-stimulus spacing (horizontal gap between 2 numbers) in the NG task and number of words in the RR task, yield different set of challenges for handling the noise.

In the RR task, initially a list of 6 words is shown to the participants and they are instructed to memorize the words. Next, a new list of words is shown to them. Some of the words from the first list are also present in the second list. The participants are asked to recognize those

Table 1. Details of the test stimulus used for the study.

	RR task	NG task
Task	(i) A list of 6 words is shown first for 30 seconds (ii) Next, a new list of words is shown (iii) Identify words from the new list those match with the first list (iv) Gaze at the word for 2 seconds and then click on that word (v) Repeat the steps (iii) and (iv) for all such matching words	(i) A sequence of 4 numbers (pin) is given (e.g., '1234') beforehand (ii) A sequence of 9 numbers is shown on the screen (iii) Gaze each number on the screen in the sequence as given in the pin (iv) Click on the number after gazing for seconds (v) Repeat the steps (iii) and (iv) for each number in the pin
Entities	Words taken from NIMHANS neuropsychological test battery [45]	Numbers (1-9)
Arrangement of entities	Multiple rows, 2 columns	Single row, 9 columns
Variations in the task	Difference in number of words in a column	Difference in inter-digit spacing between the 9 numbers with FOVs 3.28°, 2.29° and 1.642° (100, 70 and 50 pixels), respectively
Font size	48 pixels	Font sizes used are 50, 35 and 25. The ratio of inter-digit spacing and font size is kept constant to 2:1

<https://doi.org/10.1371/journal.pone.0196348.t001>

words and click on them. The font size of the words is selected to be 48 pixels (1.57°) [46] and the words are presented in a 2 column format. In order to evaluate the performance of noise cleaning algorithms (gaze tracking), the number of words per column is varied from 6 to 16. Fig 3 shows a snapshot of the designed task. The words have been chosen from National Institute of Mental Health and Neurosciences (NIMHANS) neuropsychological test battery [45].

The second test stimulus consists of a NG task wherein, a lookup table of 9 digits (1 to 9) is shown at the top of the screen as shown in Fig 4. The participants are instructed to gaze at 4 digits, one after the other, in a predefined sequence (as communicated by the instructor/experimenter) before starting the task. The inter-number spacing (*S*), and the font size (*f*) are varied in each trial keeping the *S/f* ratio constant. Three trials are conducted with inter-digit spacing

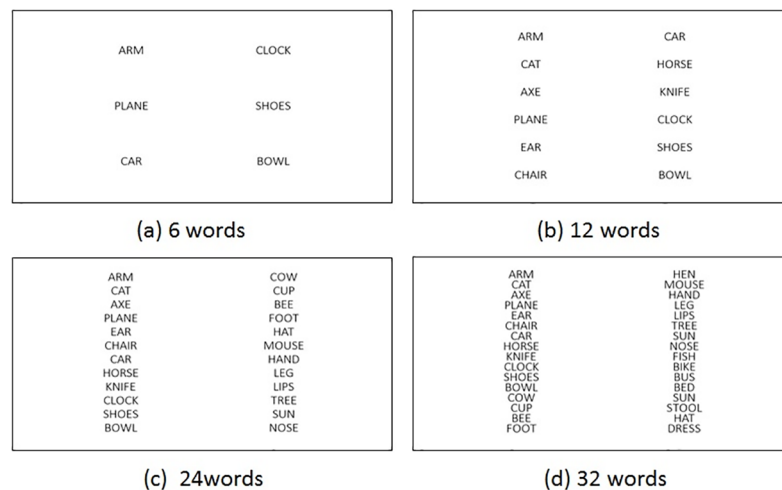


Fig 3. Designed recall-recognition (RR) test (a) 6 words for recall (3 words/column); and (b) 12 words (6 words/column) (c) 24 words (12 words/column) and (d) 32 words (16 words/column) for recognition.

<https://doi.org/10.1371/journal.pone.0196348.g003>

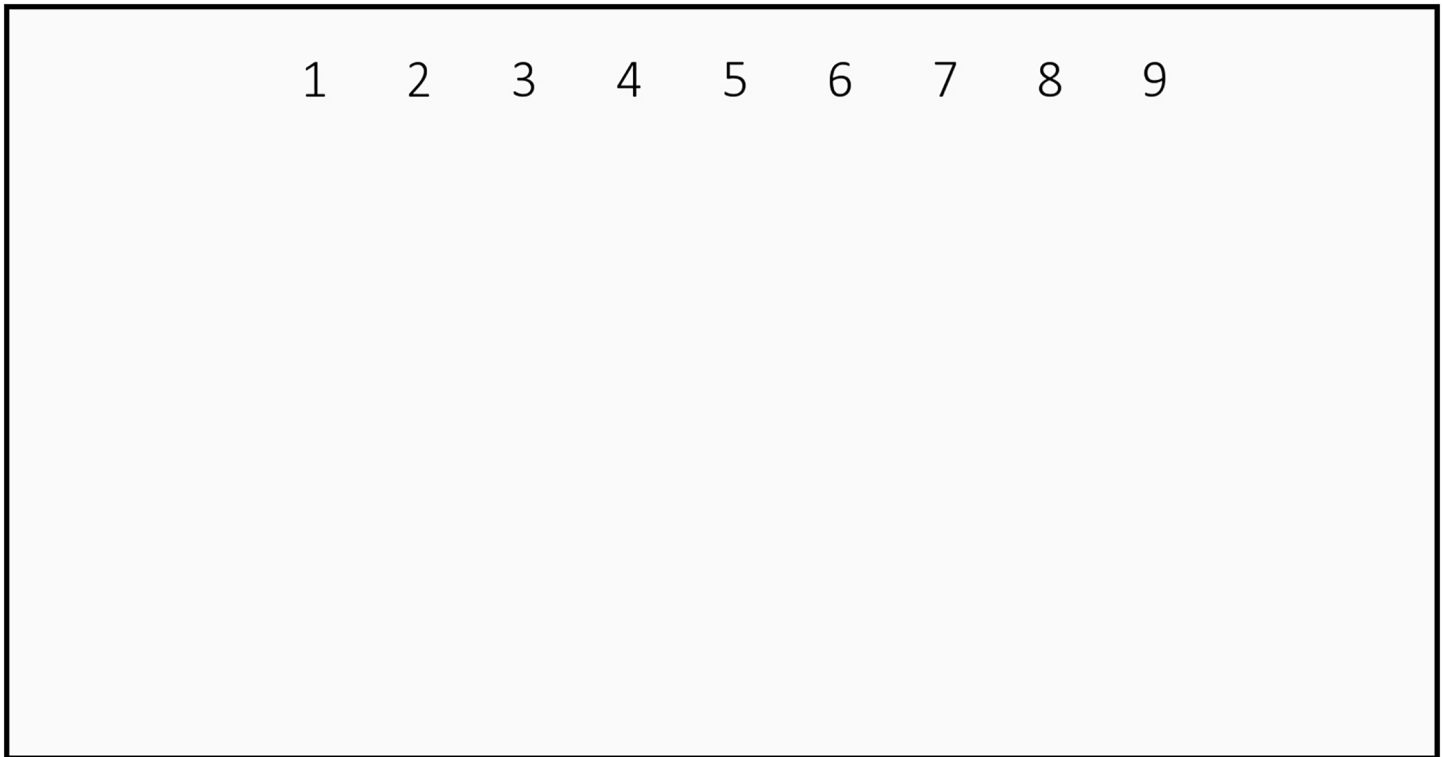


Fig 4. Number gazing (NG) task for inter-digit spacing of 100 pixels.

<https://doi.org/10.1371/journal.pone.0196348.g004>

of 100, 70 and 50 pixels, respectively. The font size f for the 3 different spacings are respectively, 1.675° , 1.182° and 0.854° .

After gazing at a particular word or a digit, the participant is instructed to click on the same. The coordinates and timestamps of the click event are logged in order to segment the corresponding eye gaze data for further analysis.

The underlying motivation for the selection of these test stimuli are as follows. Recently various medical applications are being developed keeping in mind the overall mental/cognitive well-being. On the other hand, alternative communication aids for patients with neuro-motor disorders, controlling wheel chair or various devices through human computer interfaces [47], gaze tracking-based applications for patients suffering from Autism [48], and also various standard cognitive assessments like SDMT [49], trail making task, etc. are some of those kinds of applications. In most of these applications, attention, memory retention, working memory, etc. are important aspects. RR test used in the present study is a standard psychological test, which is used to assess higher order cognitive functions, like memory retention capacity and attention. On the other hand, the NG task additionally involves the usage of working memory in order to correctly sequence the fixations on the given order of numbers. Thus the stimuli used, closely resembles the tasks that are performed by psychologists to assess cognitive functionalities of an individual.

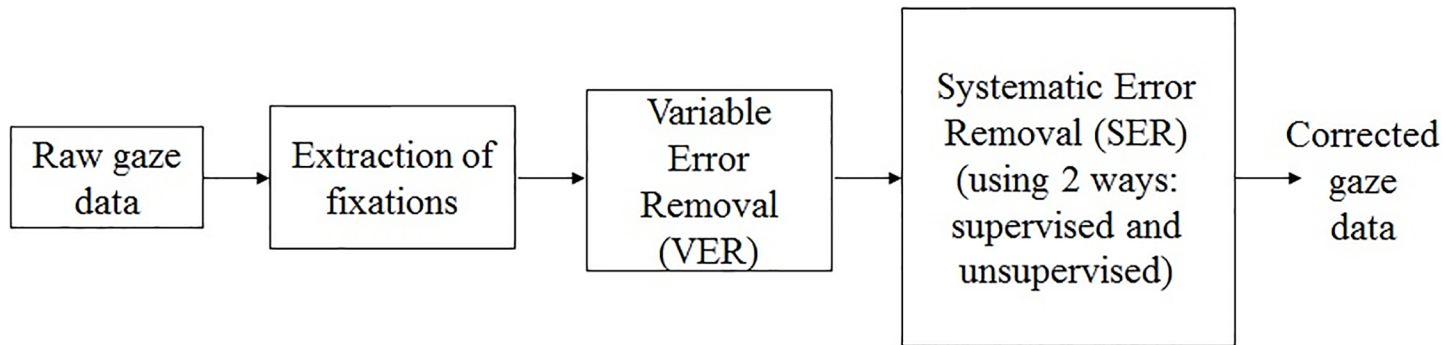


Fig 5. Block-diagram of the proposed method.

<https://doi.org/10.1371/journal.pone.0196348.g005>

4 Methodology

In the present study, we propose a novel approach of handling errors by using the nature of the drift or the systematic error from the neighborhood regions. Along with this, we have applied the principle of ‘ n -nearest stimulus points’ and have adopted unsupervised techniques. In addition to this, a novel supervised method based on the concept of ‘ n -nearest calibration points’ is applied. The essence of ‘ n -nearest’ over ‘closest stimulus point’ [14] and ‘ n -nearest’ over ‘closest calibration point’ is introduced along with inverse weighing function-based approximation. The accuracies of all these approaches are compared.

This section explains the noise removal approaches adopted in the present study. It also explains the methodology adopted for generating corrected gaze data from raw eye tracker data. The overall process is shown in the Fig 5.

First the fixation data are extracted from the raw (unprocessed) eye tracker data. Next, the variable error is filtered from this data. Finally, the systematic error is removed to get corrected gaze coordinates using both supervised and unsupervised approaches. Each of these approaches are explained in detail in Figs 6 and 7.

4.1 Extraction of fixations from raw eye tracker data

Eye movement data can be classified into 2 major classes, i.e., fixations and saccades. Many works pertaining to the classification of eye tracker data into these classes exists [50], [51, 52]. Enkelejda, et al [51] proposed the usage of low resolution eye tracker in approximating the clusters of fixation to a region of interest (ROI) using online bayesian learning. However, even in a given fixation chunk, the variable error persists. We have extracted the fixations from the raw eye tracker data using the velocity threshold-based method (VT) as explained in [50]. Eye gaze data usually consists of fixations and saccades. The data points lying above a threshold velocity are treated as saccades and the rest are categorized as fixations. As suggested in [53], we have used the velocity threshold value as 20.

4.2 Variable error removal (VER)

To handle the fluctuations or the variable error in the fixation data, we have made a survey of various filtering approaches available in literature. In the present study, we have used the graph signal processing (GSP) [54] and Kalman filter (KF) [55] for removing the variable error. The application of GSP and KF on the raw eye gaze data is explained as follows:

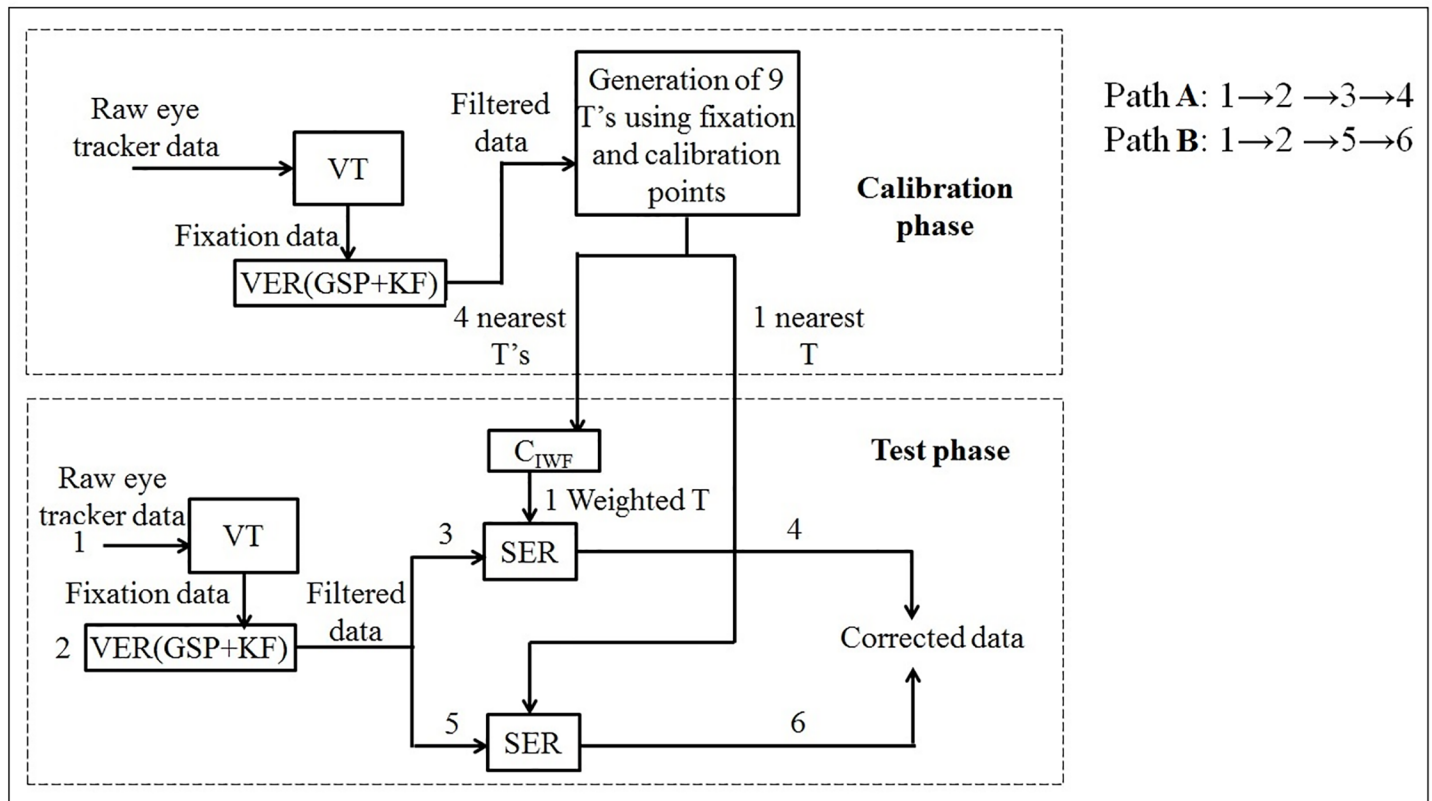


Fig 6. Supervised approach for data correction.

<https://doi.org/10.1371/journal.pone.0196348.g006>

4.2.1 Graph signal processing-based signal cleaning. During the data capture, the eye tracker captures the noisy (x, y) gaze coordinates on a plane (monitor) reported by the eye tracker and can be represented as Eq (1),

$$S = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]^T \quad (1)$$

where n is the number of samples in the signal. In our experiment as we focus on a single position on the screen, the eye gaze signal should return a single coordinate. However, S fluctuates due to the presence of variable error. Sometimes, the fluctuations are not mere oscillations around the actual position, rather these are far away from it. Hence, the denoising algorithm needs to be designed in such a way that it can handle those abrupt changes and produce a smooth signal, which is close to the actual eyeball location. Hence, GSP is suitable for this application as it smoothes the signal in accordance with the underlying graph structure, unlike other low pass filtering (LF) methods [31]. In order to perform the denoising, S is divided into a number of non-overlapping windows of length $L (\leq n)$ and then GSP-based denoising is applied on each of these windows separately. In order to do so, first a graph signal $G(V, E, A)$ is formed, which is characterized by a set of vertices V , set of edges E and an adjacency matrix A , which stores the weighted connection between the vertices. In our case, V is formed by taking the coordinates (x_i, y_i) in a particular window. The connection is formed by keeping all the vertices pairs between which an edge exists. The edges are formed if the

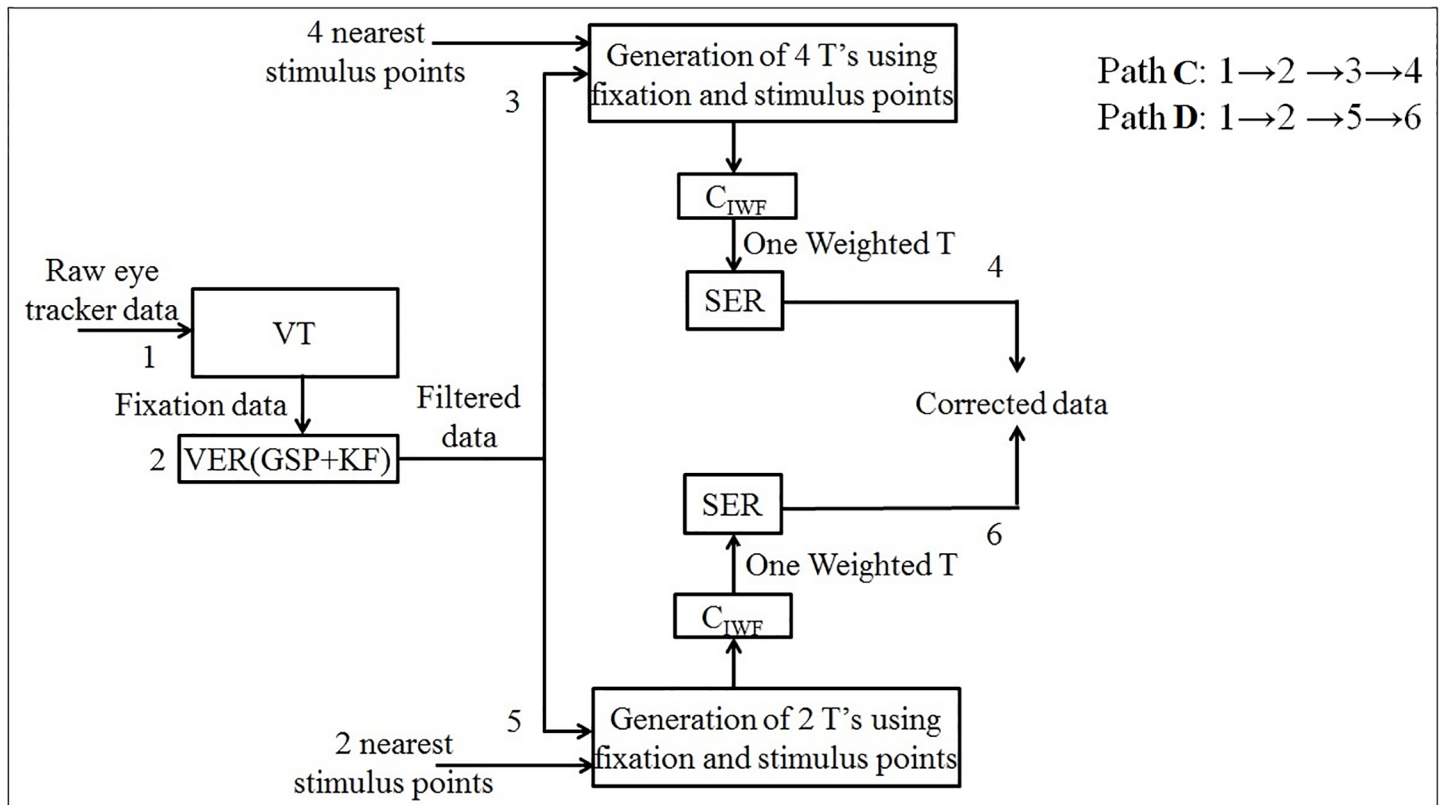


Fig 7. Unsupervised approach for data correction.

<https://doi.org/10.1371/journal.pone.0196348.g007>

Euclidean distance between the two vertices n and m is less than a threshold value th (empirically taken as 1) and the set of edges can be expressed as,

$$E = (n_i, m_j) \tag{2}$$

In this type of range-based searching, each of the vertices has different number of neighbors in a particular window, which introduces dynamicity in the graph formation and provides an edge over other filtering methods. The weighted adjacency matrix A is constructed by putting weights on edges depending on closeness measure between the two vertices. Closeness between two vertices is measured by the Euclidean distance between those 2 vertices. Hence, the weight of the connection between vertices n and m is defined using a Gaussian kernel for a constant θ as shown in Eq (3),

$$\vec{C}_B = \begin{cases} \exp\left(\frac{-|dist[n,m]|}{\theta^2}\right), & \text{if } (n, m) \in E \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

In our study, θ is chosen to be 1. The graph signal G formed in each window is corrupted by variable noise and can be written as,

$$G = t + e \tag{4}$$

where t is the clean signal and e is the noise added to it. In order to obtain the clean signal that is close to the original signal as well as smooth, a multi-objective optimization can be formed in a quadratic form as,

$$\operatorname{argmin}_t \left(\frac{1}{2} \|t - G\|^2 + \alpha \frac{1}{2} \|t - At\|^2 \right) \tag{5}$$

Here, α controls the amount of smoothness desired in the estimated S_t . This stated optimization problem can be solved by setting the first derivative of Eq (5) to zero and the closed form solution can be derived as,

$$\frac{1}{2} \frac{\partial}{\partial t} (\|t - G\|^2 + \alpha \|t - At\|^2) = 0 \tag{6}$$

$$\frac{1}{2} \frac{\partial}{\partial t} ((t - G)^*(t - G) + \alpha t^*(I - A)^*(I - A)t) = 0 \tag{7}$$

$$S_t = (I + \alpha(I - A)^*(I - A))^{-1}G \tag{8}$$

where $*$ is Hermitian of the matrix. The solution stated in Eq (6) denoises the graph signal in each window as shown in Eq (8). The formation of the graph is dependent on the size of the window, which can be chosen judiciously. A bigger window provides a smoother signal which is more influenced by the abrupt fluctuations present in that window, whereas smaller windows fail to smooth the signal efficiently. Here, we have heuristically taken L as 10. The most expensive step in Eq (8) is the inversion of the matrix $(I - A)$. In our case the size of A is only of 10×10 , which makes the $(I - A)$ inversion affordable in terms of computation. The pseudocode is provided in Algorithm 1.

Algorithm 1: Pseudocode for Graph Signal Processing based signal cleaning

```
GSPfiltering( $S_x, S_y$ ) timeseries of eye gaze position;
Input : Timeseries of eye gaze coordinates  $S$ 
Output : GSP filtered eye gaze data  $S_t$ 
Initialization : Window length  $L = 10$ ,  $\theta = 1$ ,  $\alpha = 5$ ,  $th = 1$ 
FOR each time window  $k$ 
  Graph  $G(V, E, A)$  Formation :
    Edge  $E$  formation by finding the eye gaze positions whose
    euclidean distances fall inside the unit circle ( $th = 1$ )
    Compute the closeness measure as euclidean measure between two
    connected vertices
    Edge weight  $C_B$  is defined as a Gaussian kernel over closeness
    measure with constant  $\theta$  if there is any edge between two verti-
    ces or 0 otherwise
    Adjacency matrix  $A$  is computed for  $G(V, E, A)$ 
  Estimation of clean signal ( $S_t$ ) for a window:
    Filtered signal,  $S_t = (I + \alpha(I - A)^*(I - A))^{-1}G$ 
  Loop continue for other windows
```

4.2.2 Kalman filter (KF)-based signal enhancement. In order to minimize the noise further, we have used KF on the GSP filtered data $S_t\{S_x, S_y\}$. The state vectors at time k is given by, $\vec{R}_k = [S_{x_k} S_{y_k} \dot{S}_{x_k} \dot{S}_{y_k}]$, where, $\dot{S}_{x_k} \dot{S}_{y_k}$ denotes the velocity of eyeball among the X and the Y directions, respectively. The instantaneous eye movements depend on the prior velocities, i.e. $\vec{v}_j = \{S_{x_j} S_{y_j}\}$ where time ($j < k$) and hence we have modeled \vec{v}_k as the weighted sum of previous

velocities. The dynamic equations that govern the position $\vec{p}_k(S_{x_k}, S_{y_k})$, for (x, y) position at instance k , of eye gaze data are,

$$\vec{p}_k = \vec{p}_{k-1} + T * \vec{v}_{k-1} \tag{9}$$

$$\vec{v}_k = a_{k-1}\vec{v}_{k-1} + a_{k-2}\vec{v}_{k-2} + a_{k-3}\vec{v}_{k-3} + \phi \tag{10}$$

where $T = \frac{1}{f_s}$; f_s is the sampling frequency (30 Hz) of the Eye Tribe eye tracker. It is observed that the eyeball velocity follows ARIMA (3, 0, 0) or AR(3) [56] and hence, we have derived the coefficients a_{k-1} , a_{k-2} , a_{k-3} and ϕ from the ARIMA model. It is required that the coefficients are to be derived separately for each participant. The discrete state space model for eye gaze data is given by the linear stochastic difference at time k as,

$$\vec{R}_k = F\vec{R}_{k-1} + \vec{w}_{k-1} \tag{11}$$

$$\vec{Z}_k = H\vec{R}_k + \vec{\rho}_k \tag{12}$$

where F is the state transformation matrix. The actual observation is made at time k . The noiseless connection among the measurement vector \vec{Z}_k and state vector \vec{R}_k is designated by H . The $\vec{\rho}_k$ and \vec{w}_k are measurement and process noise (uncorrelated gaussian noise following zero mean and co-variance of ϕ_2 and ϕ_1), respectively. The Kalman filter corrects the eye gaze data \vec{R}_k after receiving \vec{Z}_k (at time k) by,

$$\hat{R}_k = \vec{R}_k + K_k(\vec{Z}_k - H\vec{R}_k) \tag{13}$$

where K_k is the Kalman gain [57] and \hat{R}_k is the filtered data at time k . The pseudocode is as provided in Algorithm 2.

Algorithm 2: Pseudocode for Kalman filtering for denoising the signal

```

KalmanSmoothing ( $S_{t_x}, S_{t_y}$ );
Input : GSP filtered eye gaze coordinates  $S_t = (S_{t_x}, S_{t_y})$ 
Output : Smooth data  $\hat{R}_k$ 
Initialization: Estimated state vector  $\vec{R}_{k|k-1}^+$ , state transition matrix  $F$ ,
measurement mapping matrix  $H$ , process noise co-vari-
ance matrix  $\phi_1$ , measurement noise co-variance matrix
 $\phi_2$ , priori state co-variance  $\vec{P}_{k-1}$ 
FOR each time epoch  $k$ 
  Prediction : State prediction based on  $(k - 1)^{th}$  state given  $F$  and  $\phi_1$ 
  Update : Update the posterior mean of state estimate based on
the new measurement  $\vec{S}_t^k$  given  $\phi_2$  and  $H$ . Compute Kalman
Gain  $K_k$  and update the covariance  $\vec{P}_{k-1}$  and state esti-
mate  $\hat{R}_k$ 
Loop continue
    
```

4.3 Systematic error removal (SER) using linear transformation (LT)

The filtered data, \hat{R} is subjected to further processing with spatial transformation in order to remove the systematic error [39]. The method basically uses the separation between the actual and the desired (ground truth) gaze coordinates for generating a 2×2 transformation matrix

T . The best-fitting values of T are obtained using optimization routines such as simplex algorithm [39]. Next, this matrix is used to correct the actual fixation data \hat{R} as,

$$\vec{C} = \hat{R} * T \tag{14}$$

where \vec{C} is the corrected gaze data. In ideal case, if \hat{R} exactly matches the ground truth coordinates, T would have been an identity matrix. We have applied both supervised and unsupervised approaches for error removal.

In the supervised approach, the systematic error is learnt in the calibration phase in terms of transformation matrix T , which is then used in the succeeding test phase (for supervised approach), whereas it is derived directly from the test data set in case of unsupervised approaches, discussed in the following subsections.

4.3.1 Supervised approaches—Paths A and B. The designed supervised approach is depicted in Fig 6.

The fixation data is extracted from the eye gaze data collected in the calibration phase, as explained in section 4.1. Next, the data is subjected to variable error removal as explained in section 4.2. Finally, the transformation matrix T is derived for each of the 9 static points (S1 through S9). Each of the T s are evaluated for correctness. We define a correctness measure M , given by,

$$M = \det(T) \tag{15}$$

Ideally, if the raw data and the ground truth data exactly matches, T would be an identity matrix with $M = 1$. We computed M for over 200 fixation chunks and a threshold of 0.8 is set empirically. If $M < 0.8$ for any calibration point, then it is rejected and is replaced by the average T of 2 nearest calibration points whose M value is greater than the threshold. If more than 3 T s have M less than threshold of 0.8, a fresh set of data are captured for the calculation of transformation matrices for that particular participant.

The matrices derived from the calibration phase are stored and are used to remove the systematic error in the test phase. The proposed method deals with extracting the transformation matrix T for a given fixation chunk with centroid X from its nearest calibration point S . The main principle behind this assumption is that the nature of systematic error for the given fixation chunk is similar to the systematic error seen on the nearest calibration point S (during the calibration phase), computed using k-nearest neighbor search algorithm [58]. Path A, as shown in Fig 6 has 4 T s based on inverse weighing function \vec{C}_{IWF} , defined as,

$$\vec{C}_{IWF} = \begin{cases} \frac{\sum_{i=1}^N w_i(x) \vec{u}_i}{\sum_{i=1}^N w_i(x)}, & \text{if } d(x, y) \neq 0 \forall i \\ \vec{u}_i, & \text{otherwise} \end{cases} \tag{16}$$

where $\vec{u}_i = \hat{R}_k * T_k$ for $k = 1$ to 4 nearest calibration points; $N = 4$ nearest neighbors; $w_i(x) = \frac{1}{d(x,y)^p}$ where, d is the Euclidean distance between the centroids of the fixation data x and the calibration point (any one among the points S1 through S9) and the value of p is set to 2. The weight w is normalized by dividing each of the 4 weights by the sum of the total weight. The weights are inversely proportional to the square of the distance, which implies that the corrected gaze data is mostly influenced by the nearer neighbors. Nearer the point, more the force applied to pull the point towards the calibration point. For path B, as shown in Fig 6, only the transformation matrix corresponding to the most nearest calibration point is taken into account for correcting the systematic error. Hence, in the current supervised approach, we

have used either 4 and 1 nearest calibration points. The reason behind using 4 points is to check the influence of systematic error across the screen (i.e. in terms of magnitude and direction of the drift). In contrast to this, we have checked the performance of the filtered fixation data against 1 closest calibration point, which incorporates the nature of the systematic error corresponding to that particular point only. The pseudocodes for paths A and B are provided in Algorithms 3 and 4, respectively.

Algorithm 3: Pseudocode for Path A

Input : Raw Gaze data, $S = (S_x, S_y)$
Output : Variable and Systematic error removed data, C
Procedure: S = Extract fixation data from raw eye gaze data, S
Variable Error Correction:
 $S_t = \text{GSPfiltering}(S_x, S_y)$
 $\hat{R}_k = \text{KalmanSmoothing}(S_{t_x}, S_{t_y})$
Systematic error removal:
Obtain 1 T derived from 1 nearest calibration point
Obtain corrected data C, by transforming the \hat{R}_k using T as,
 $C = \hat{R}_k * T$

Algorithm 4: Pseudocode for Path B

Input : Raw Gaze data, $S = (S_x, S_y)$
Output : Variable and Systematic error removed data, C
Procedure: S = Extract fixation data from raw eye gaze data, S
Variable Error Correction:
 $S_t = \text{GSPfiltering}(S_x, S_y)$
 $\hat{R}_k = \text{KalmanSmoothing}(S_{t_x}, S_{t_y})$
Systematic error removal:
Obtain 4 Ts derived from 4 nearest calibration points
Find weighted T from the 4 Ts using inverse weighing function
Obtain corrected data C, by transforming the \hat{R}_k using T as,
 $C = \hat{R}_k * T$

4.3.2 Unsupervised approach—Paths C and D. This approach is based on the ‘n—nearest stimulus point’ (not the calibration point), in contrast to the ‘required fixation location’ [12], ‘probable fixation location’ [39], and ‘closest stimulus point’ [14]. In our case, $n = 2$ or 4 neighboring stimulus points.

For path C, as shown in Fig 7, 4 nearest stimulus points are selected and transformation matrices with respect to each of these 4 locations are derived. Later, inverse weighing function is applied, as discussed in Eq (16), to get the corrected data. For path D, as shown in Fig 7, similar approach is applied but with only 2 nearest stimulus points instead of 4 (Fig 8). For the fixation data represented by black dots, the 4 nearest stimuli are A, F, P and X with the Euclidean distances d_1, d_2, d_3 and d_4 from the fixation center. The weights are chosen to be inversely proportional to the distance, i.e. lesser the distance, larger the weight; which implies that the corrected fixation would be more biased towards the nearer neighbors. In the unsupervised approach, we have used 4 and 2 nearest stimulus points. The usage of 4 points is analogous to the one mentioned in supervised approach. The pseudocodes for paths C and D are provided in Algorithms 5 and 6, respectively.

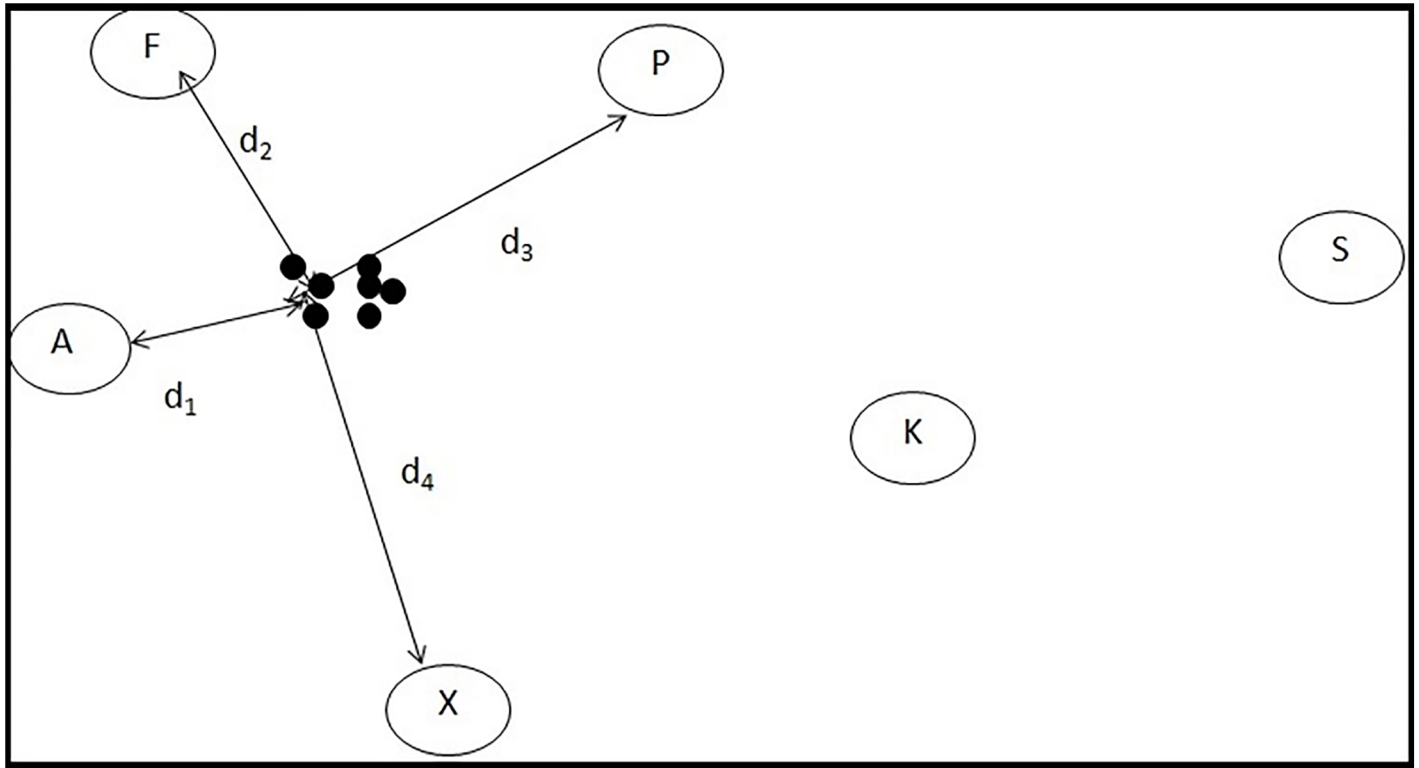


Fig 8. Demonstration of inverse weighing function for 4 nearest neighbor stimuli points.

<https://doi.org/10.1371/journal.pone.0196348.g008>

Algorithm 5: Pseudocode for Path C

Input : Raw Gaze data, $S = (S_x, S_y)$
Output : Variable and Systematic error removed data, C
Procedure: S = Extract fixation data from raw eye gaze data, S
 Variable Error Correction:
 $S_t = \text{GSPfiltering}(S_x, S_y)$
 $\hat{R}_k = \text{KalmanSmoothing}(S_{t_x}, S_{t_y})$
 Systematic error removal:
 Obtain 4 T s derived from 4 nearest stimulus points
 Find weighted T from the 4 T s using inverse weighing function
 Obtain corrected data C, by transforming the \hat{R}_k using T as,
 $C = \hat{R}_k * T$

Algorithm 6: Pseudocode for Path D

Input : Raw Gaze data, $S = (S_x, S_y)$
Output : Variable and Systematic error removed data, C
Procedure: S = Extract fixation data from raw eye gaze data, S
 Variable Error Correction:
 $S_t = \text{GSPfiltering}(S_x, S_y)$
 $\hat{R}_k = \text{KalmanSmoothing}(S_{t_x}, S_{t_y})$

Systematic error removal:

Obtain 2 T s derived from 2 nearest stimulus points

Find weighted T from the 2 T s using inverse weighing function

Obtain corrected data C , by transforming the \hat{R}_k using T as,

$$C = \hat{R}_k * T$$

From the above discussion it is clear that unsupervised approaches mainly aim on dragging the gaze data towards its nearest stimulus, whereas, supervised approaches handle the gaze data by considering the direction and magnitude of the systematic error as obtained in the calibration phase. However, usage of any one of this method solely cannot serve all the types of stimulus. For instance, if the stimulus points are very densely packed (e.g. designed stimulus Recall-Recognition (RR) task with more than 24 words), then the nearest stimulus-based noise cleaning fails considerably. In such cases, the supervised approaches can prove to be beneficial. In contrast to this, in case of stimulus points being placed far apart (e.g. designed stimulus Number Gazing (NG) task), the gaze data could be handled very well using the nearest stimulus positions. Hence, the accuracy of noise cleaning relies on the nature of the stimulus and so we have experimented with both the supervised and unsupervised approaches.

It is to be noted that the four paths A, B, C and D are independent of each other and we tested them one after the other on the data to check the effectiveness of each of them.

5 Experimental paradigm

This section discusses the experimental setup, various stimuli used and the details of the data collection procedure adopted in the present study.

5.1 Setup

The experimental setup is shown in Fig 9. We have used a low-cost eye tracker from Eye Tribe [59] having a sampling rate of 30 Hz. The Eye Tribe device is placed below the screen as shown in Fig 9. An wooden chinrest fixed on the table is used while collecting the eye gaze data. A height adjustable chair was used during data collection. The stimulus is shown on a computer screen (1366 × 768) placed at a distance of approximately 60 cm from the participants. The entire experiment is carried out in a closed, quite room under constant lighting conditions.

5.2 Participants

Twenty participants (12 females and 8 males, mean age 32 ± 5.54 years) are selected from our research lab for the experimentation. All of them had normal or corrected to normal vision with spectacles. We have ensured that they belong to similar cultural backgrounds and have similar educational qualification. None of them had any background history of any mental or physical ailments. Participants are allowed to wear spectacle during data collection. The clearance on ethical issues for handling and analysis of the data collected has been acquired from Institutional Review Board of Tata Consultancy Services Ltd. (TCS). Informed consent is also taken from the participants and the data is anonymized for further processing.

5.3 Data collection

The experimental protocol and the tasks to be performed are explained to the participants before starting the experiment. The participants are asked to sit comfortably on a chair with adjustable height and a chin-rest is used to minimize the head movements. The participants then performed an initial software development kit (SDK)-based calibration (provided by the Eye Tribe sensor). The goodness of calibration is represented on a scale of 1-5. A score of 5



Fig 9. Experimental setup with the eye tracker at the bottom of the display and a chin rest.

<https://doi.org/10.1371/journal.pone.0196348.g009>

corresponds to best calibration giving an error below 0.5 degree, whereas the error is more than 1.5 degrees for the score of 2. Score 1 indicates the calibration is extremely bad and re-calibration needs to be performed. Before starting the actual experiment, the participants are encouraged to take part in a practice session in order to have a better understanding of the stimulus and the task to be performed. The stimulus used for practice sessions are similar to that used for the actual experimentation, however, not exactly the same in terms of the set of words to be recalled or the numbers to be gazed. The data collection is carried out in 2 phases as described in the following subsections.

5.3.1 Phase one: Initial calibration. An initial calibration (both SDK-based and our designed one) is performed once for a single (first) participant and the calibration results are applied on the remaining participants. The position of the chin rest and the eye tracker are not altered for the remaining participants. Next participant onwards, 2 tests (NG and the RR task)

are performed as explained earlier. Corresponding eye gaze data are collected and are used for further analysis.

5.3.2 Phase two: Repeated calibration. The accuracy of the protocol is later tested for ‘multiple time’ or ‘repeated calibration’, against one time calibration. Here, every participant performs both the SDK and the designed calibration, before every test session.

Participants signed a consent form before the commencement of the experimentation. The data corresponding to every participant is anonymized.

6 Results and discussion

The performance evaluation of our proposed methods are done in terms of algorithm/ approaches for the following scenarios

1. variable error removal techniques
2. comparison of supervised and unsupervised approaches for systematic error removal
3. comparison of single calibration against multiple calibration protocols
4. evaluation of proposed noise removal method for long duration tasks

We have also compared our designed approaches with the closely related state of the art methods as explained in the tree diagram given in Fig 10.

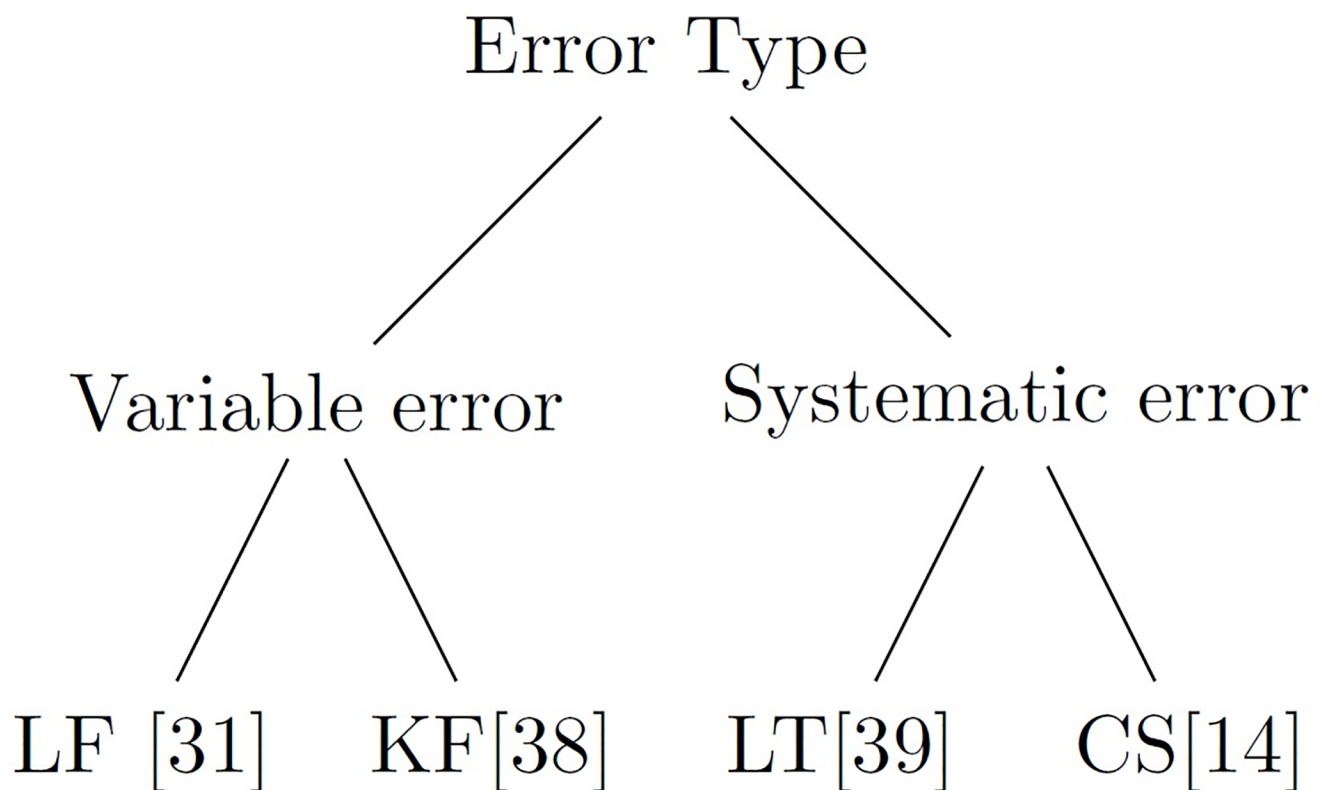


Fig 10. State of the art methods considered for comparison for different types of error, where, LF = Low pass filtering, KF = Kalman filtering, LT = Linear Transformation and CS = Closest Stimulus based approach.

<https://doi.org/10.1371/journal.pone.0196348.g010>

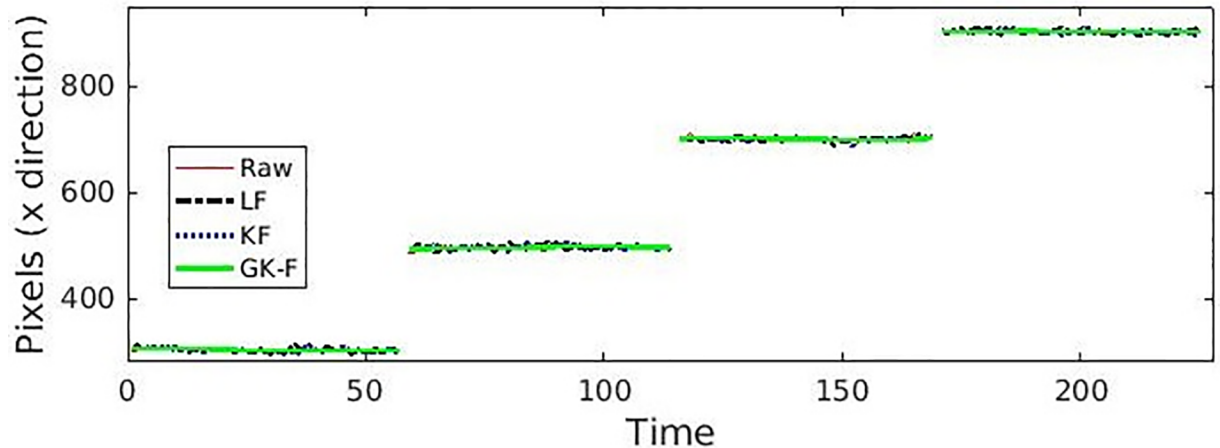


Fig 11. Comparison of different filtering approaches for the NG task wherein the participant gazed at 4 different numbers. Here, LF = Low pass filtering, KF = Kalman filtering, GSP + KF = Graph signal processing and Kalman filter.

<https://doi.org/10.1371/journal.pone.0196348.g011>

6.1 Variable error removal technique

The performance evaluation is carried out to test if our proposed method is able to extract the desired dense cluster of input fixation points.

Hence, by considering the problem associated with the variable errors, our proposed method has been compared with most widely used filtering approaches for eye tracking, i.e., Low pass filtering and Kalman Filter. Fig 11 shows the effects of different filtering approaches used on the gaze data corresponding to the NG task.

Fig 12 shows the sample results of different filters for the NG task for one particular participant (assuming systematic error to be zero). The participant is asked to gaze the numbers, 1-3-5-7. The radius of each circle reflects the general smoothness of the data. Smaller the radius of the circle, better is the filtering capabilities. For the gazed number ‘1’ in the Fig 12, the radius of raw data, low pass filter (LF) and Kalman filter (KF) filtered data is almost the same and hence, the circles are overlapping. Similarly, for the gazed number 7, the radius of Kalman filter KF filtered data is slightly larger than the GSP + KF filtered data. Note that the radii of the proposed GSP + KF data chunk are least for all the gazed numbers.

In order to get insight of the change in the radius from raw to filtered data, we have defined Smoothness Ratio (SR) metric as the ratio of the radius of raw data chunk D_r and the radius of the filtered data D_f (Eq 17). Fig 12 depicts the visualization of filtered output assuming systematic error as zero.

$$SR = \frac{D_r}{D_f} \tag{17}$$

Larger values of SR mean better the filtering approach. The estimated SR values for each filter on both the tasks are shown in Figs 13 and 14. Fig 13 shows the SR in the NG task, when the test is carried on the different categories of word spacing in comparison to the existing methods. Fig 13 clearly depicts that the performance of GSP + KF is by far better than the existing methods. There is an enhancement of 69% over the complete spacing against Raw-LF and more than 27% against Raw-KF (raw means data taken directly from the eye tracker device).

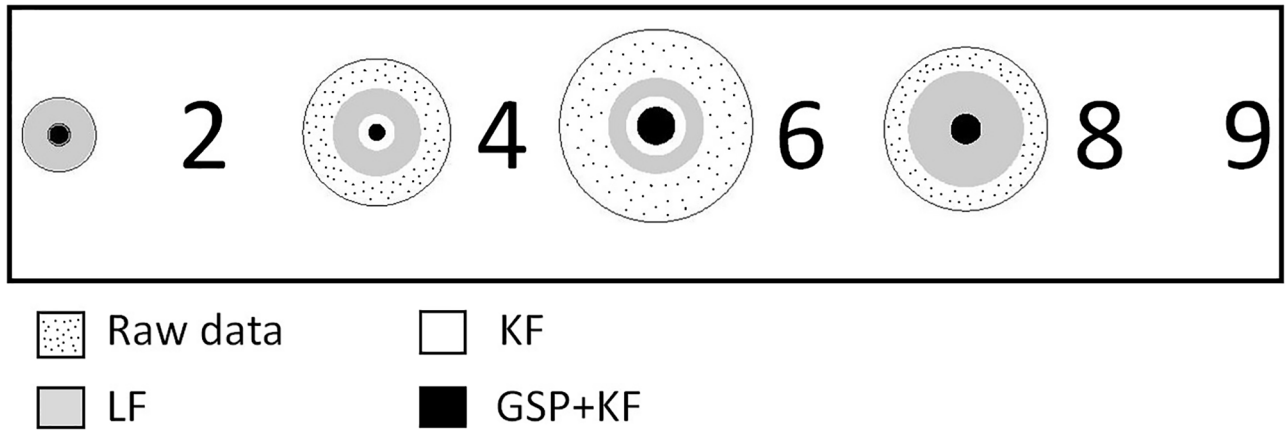


Fig 12. Demonstration of different filtering approaches in terms of smoothness for NG task. Note that for the gaze chunk on the digit ‘1’, the values of SR in terms of degrees are 0.932°, 0.92°, 0.26°, 0.2°, respectively for raw data, LF, KF and GSP + KF approaches. Here, LF = Low pass filtering, KF = Kalman filtering, GSP + KF = Graph signal processing and Kalman filter.

<https://doi.org/10.1371/journal.pone.0196348.g012>

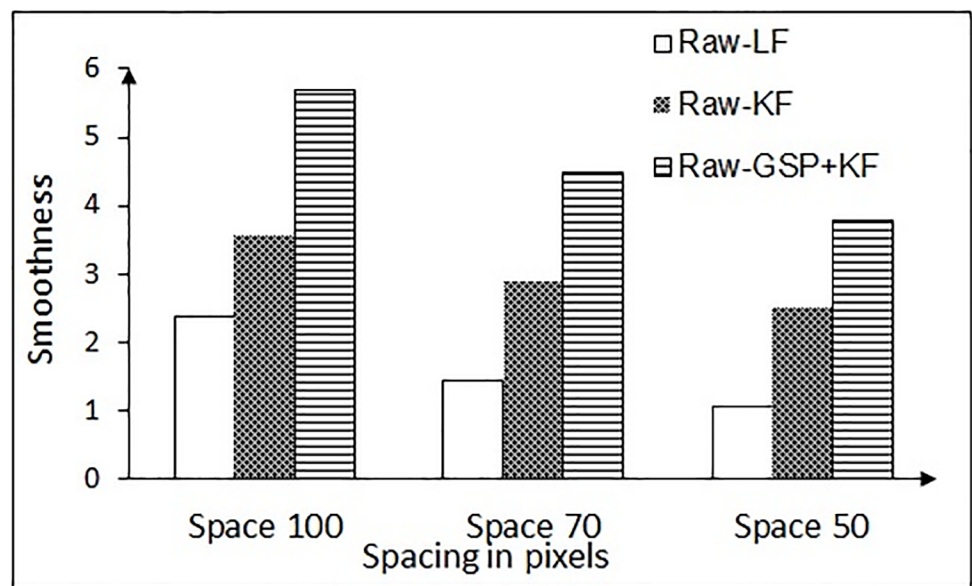


Fig 13. Smoothness ratio of proposed and existing methods in NG task. Here, LF = Low pass filtering, KF = Kalman filtering, GSP + KF = Graph signal processing and Kalman filter.

<https://doi.org/10.1371/journal.pone.0196348.g013>

SR in recall-recognition (RR) task for the proposed and existing methods is shown in Fig 14. The enhancement ranges from 56% (for minimum words) to 66% (for maximum words) compared to Raw-LF. It is to be noticed that, even though the number of words increases, the performance of the proposed method is still better.

The variable error also arises due to eye tracker hardware, exhaustion or fatigue of the user, etc. Thus a closeness measure (CL) is also required to know how the points deviate or spread across its cluster center. We have computed the CL of the data chunk with respect to its cluster

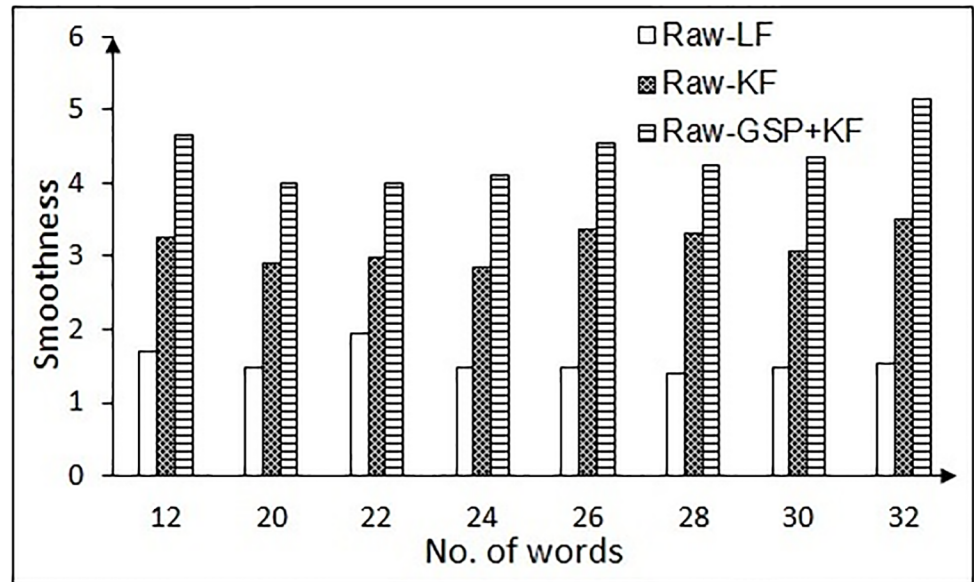


Fig 14. Smoothness ratio of proposed and existing methods in the RR task. Here, LF = Low pass filtering, KF = Kalman filtering, GSP + KF = Graph signal processing and Kalman filter.

<https://doi.org/10.1371/journal.pone.0196348.g014>

center as given by Eq (18),

$$CL = \sqrt{\frac{\sum_{i=1}^N (p^i - p_c)^2}{N}} \tag{18}$$

where N is the number of samples, $p = (x, y)$ are the fixation data coordinates and $p^i = (x^i, y^i)$ represents the coordinates of the cluster center. We call this metric as the *closeness measure*, as it computes the distance of cluster center from rest of the points. Lower closeness values indicate better filtering approach.

The results for variable error removal, based on closeness measure are presented in Figs 15 and 16, for the NG task and the RR task, respectively. It is to be noted that the combination of graph signal processing and Kalman filter performs better in comparison to the low pass filter and Kalman filter in increasing the compactness in the data chunk. Hence, this combination is used for the further analysis.

Note that in case of the NG task, the number of participants are 20, out of which for the first 10, we manually selected the chunks in the gaze data. From these chunks, the data of window length 2 seconds was taken for further analysis because as per the subjective feedback, gazing on a particular entity for more time is difficult, which would rather encompass micro-saccades. For remaining 10 participants, we introduced a ‘click’ event in the NG task and the data of window length 2 seconds before the click event was considered. This was done to verify any significant change in the performance of the algorithms due to click event and it was seen that there was no significant effect. Hence, for the rest of the stimulus types, i.e. for the NG task with multiple calibration and modified NG task for long duration analysis, the click event was used to speed up the process.

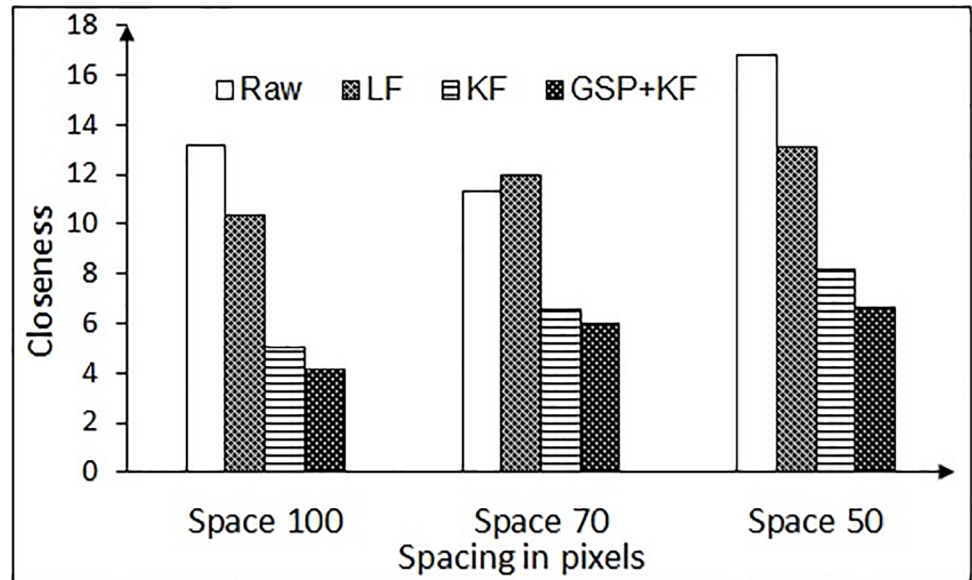


Fig 15. Closeness measure results for variable error correction of the NG task for different spacing (for one-time calibration protocol). Here, LF = Low pass filtering, KF = Kalman filtering, GSP + KF = Graph signal processing and Kalman filter.

<https://doi.org/10.1371/journal.pone.0196348.g015>

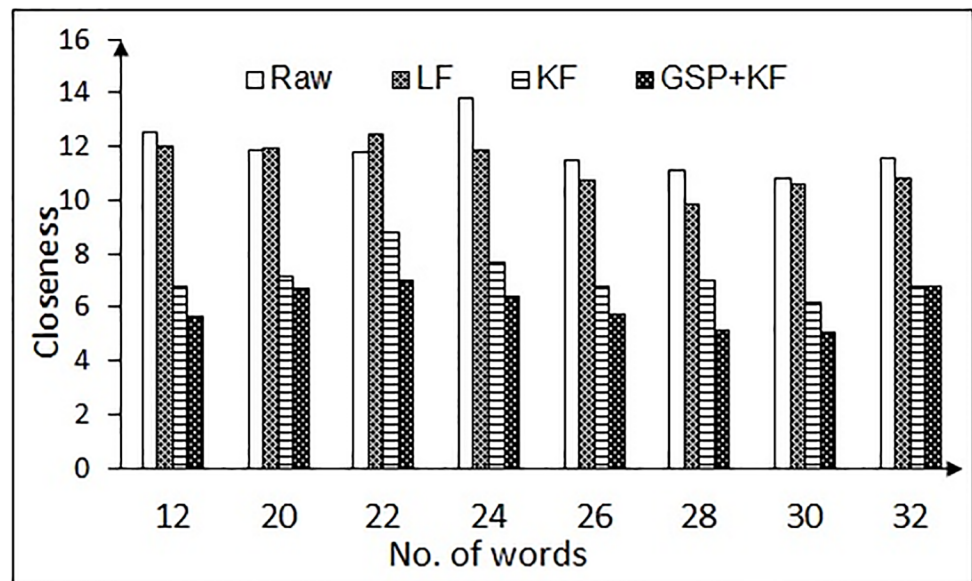


Fig 16. Closeness measure results for variable error correction of the RR task for different number of words (for one-time calibration protocol). Here, LF = Low pass filtering, KF = Kalman filtering, GSP + KF = Graph signal processing and Kalman filter.

<https://doi.org/10.1371/journal.pone.0196348.g016>

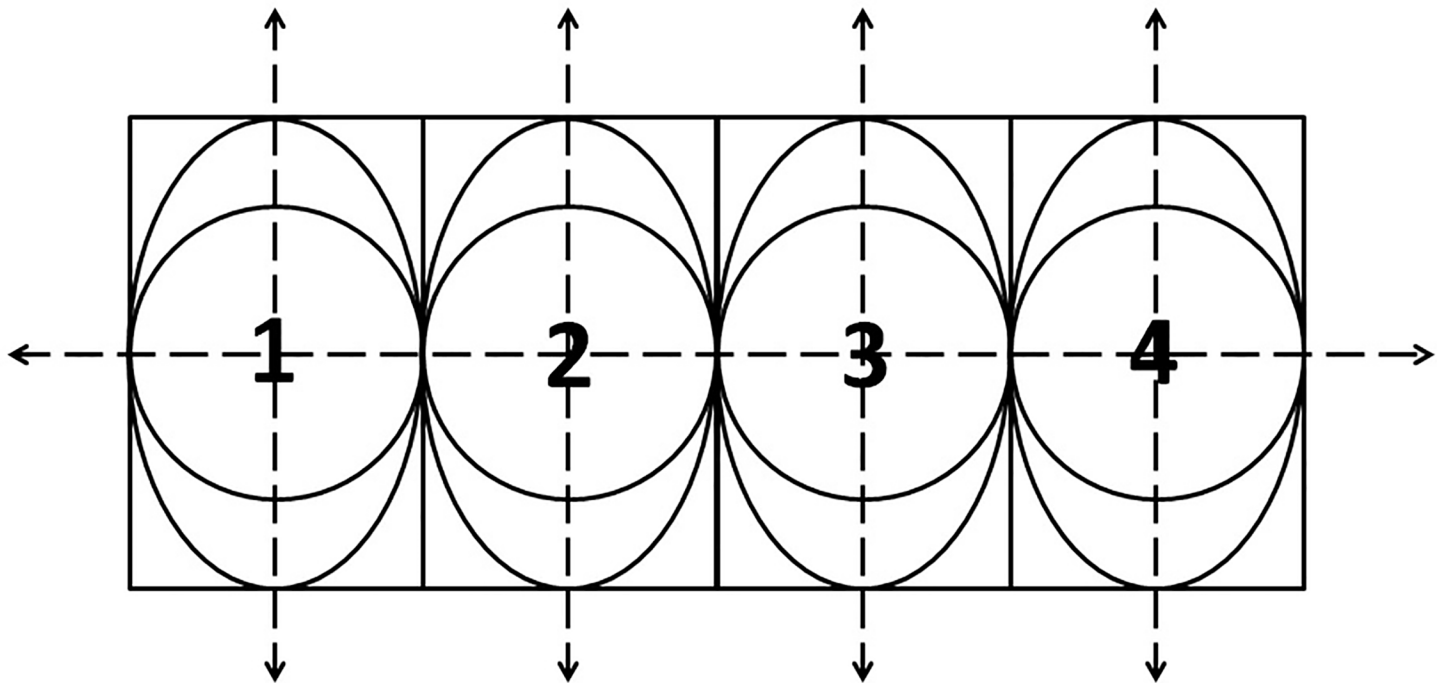


Fig 17. Pictorial scheme of different types of boundaries defined around each number in the NG task.

<https://doi.org/10.1371/journal.pone.0196348.g017>

6.2 Comparison of unsupervised and supervised approach for systematic error removal

For systematic error removal, the approach used to measure the algorithmic performance is through the accuracy of having the centroid of the gaze chunk in the area of interest (AOI) of the stimulus. Hence, efficient designing of the AOI boundary also plays a vital role in the computation of accuracy. Based on this, three different AOIs are considered, viz. circular, rectangular and elliptical, as shown in Fig 17 for the NG task. The accuracy thus obtained for raw gaze data is reported in Table 2. The reason for using the raw gaze data in this context is to throw light on the effectiveness of the boundaries in the absence of any noise cleaning techniques. It is evident from Table 2 that rectangular boundary provides maximum accuracy and hence, further analysis is carried out using the rectangular boundary only.

Next the performance of each path (Path A through D as depicted in Figs 6 and 7) is assessed in terms of accuracy for the rectangular boundary and is shown in Figs 18 and 19, for the NG task and the RR task, respectively. In case of NG task, the accuracy of the raw data decreases considerably with the decrease in inter-number spacing. However, with the

Table 2. Accuracy (%) of detecting the gazed number using the 3 different boundaries for raw gaze data.

AOI	100 Space	70 Space	50 Space
Circle	57.5	53.75	38.75
Ellipse	66.25	65	57.5
Rectangle	71.25	67.5	60

<https://doi.org/10.1371/journal.pone.0196348.t002>

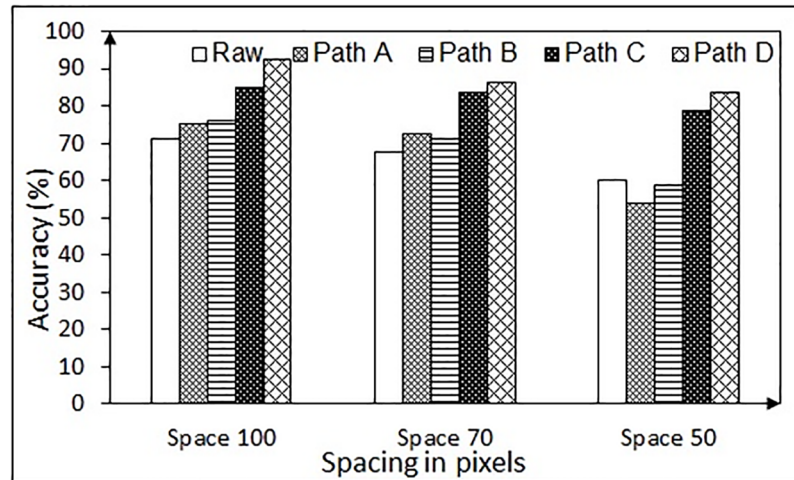


Fig 18. Accuracy of detecting the gazed numbers using different algorithmic chains for the NG task (for one-time calibration protocol).

<https://doi.org/10.1371/journal.pone.0196348.g018>

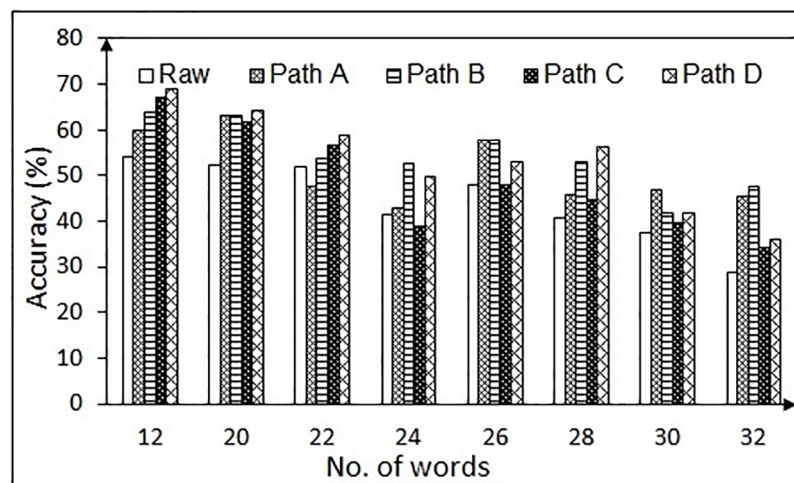


Fig 19. Accuracy of detecting the gazed words using different algorithmic chains in the RR task (for one-time calibration protocol).

<https://doi.org/10.1371/journal.pone.0196348.g019>

proposed approach, the error is reduced, thereby enhancing the accuracy. For the RR task, the accuracy of all the approaches drops with the increase in number of words (thereby decrease in spacing between the neighboring words). However, it is to be noted that it is the best possible option to use 12 words in order to get good accuracy in such systems. The results confirm that even with one-time calibration, the designed algorithmic chain can handle the variations in the gaze data due to subject-specific differences, making it a practical solution for patients who are unable to perform calibration.

Table 3. Comparison of average accuracy (%) in systematic error correction for different calibration protocol (RR-Recall recognition, NG-Number gaze).

Task	Calibration Type	Raw	Path A	Path B	Path C	Path D [39]	LT	Closest Stimulus [14]
RR	Single	28.86	45.38	47.47	34.07	36.16	18.6	38.24
	Multiple	38.33	40	38.33	36.67	38.33	10	36.67
NG	Single	60	53.75	58.75	78.75	83.75	78.75	88.75
	Multiple	70	82.5	82.5	87.5	90	82.5	100

<https://doi.org/10.1371/journal.pone.0196348.t003>

6.3 Comparison of single calibration against multiple calibration methods

We have further compared the proposed and the existing noise cleaning methods considering multiple calibration (i.e. calibration for each participant) and proposed one time calibration. The stimulus chosen for comparison are NG task with 50 pixels (inter-digit spacing) and RR task with 32 words (i.e. 16 words/column). The comparison of accuracy of handling the systematic error for single and multiple calibration is given in Table 3. The nature of systematic error is calibration-dependent, which is evident from this table. Also, with multiple calibration, the accuracy of detecting the desired stimulus point increases, however, the difference in accuracy for both the cases is not considerably high. Hence, in cases where multiple calibration is not possible, it is acceptable to go ahead with single calibration, provided the eye tracker and chin rest positions are not altered.

In Table 3 it can be seen that the overall accuracy for the supervised approach (for RR task) for multiple calibration is lesser in comparison to its single calibration counterpart; in contrast to the NG task where the reverse behavior is seen. This can be attributed to the fact that not all the participants were able to perform the calibration phase properly, thereby degrading the overall accuracy. This behavior is seen specifically for the RR task as the stimulus points (words) are densely packed in this case, hence, the effect of the n -nearest calibration points for the supervised approach has more impact which might be degraded due to bad calibration. In case of the NG task, the calibration points at the top of the screen had more impact owing to the placement of the numbers on the screen.

Table 3 shows the comparison of various proposed approaches over the closely related state-of-the-art methods. It is to be noted that in case of NG task, the proposed unsupervised approach (Path C and D of Fig 7) outperforms [39]. It is also evident from the Table 3 that though the method proposed in [14] outperforms all other methods for NG task, but it does not work in cases where the number of stimuli points is large or the stimuli are densely packed. Hence, the results obtained using [14] is not good for the RR task. The main reason is that the method in [14] is based on closest stimulus point and hence, the accuracy is computed by looking at the closest stimulus point from the gaze chunk obtained using mean shift algorithm [14]. In our case, the error free data either lies in the correct stimulus region $\{R_c \in U\}$, wrong stimulus region $\{R_w \in U\}$ or in no-man's land $\{R_n \in U, \text{ where } R_n \notin (R_c \cup R_w)\}$; due to the rectangular boundary defined around each stimulus point, where U corresponds to the overall screen region. However, it is to be noted that the method given in [14], forcibly moves a fixation to the closest point and hence, a true fixation away from the stimulus is not detected by the algorithm. Therefore, all our proposed methods are performing better than existing methods as reported in Table 3, maximum accuracy is obtained with Path B for RR task and that for NG task is obtained through Path D.

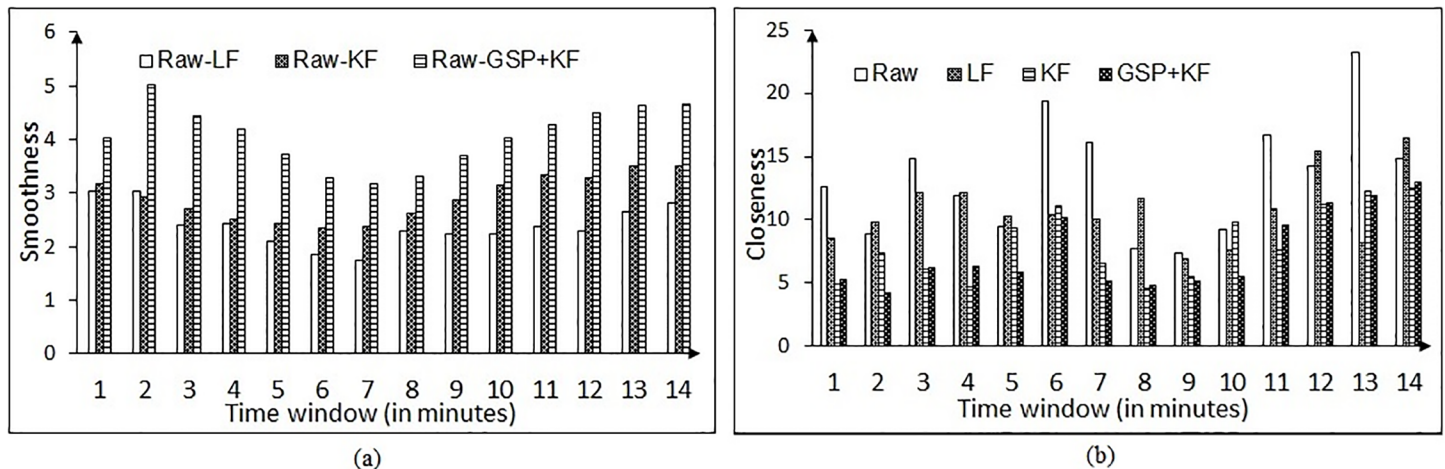


Fig 20. Removal of variable error for participant P1. (a) Smoothness parameter, (b) Closeness parameter.

<https://doi.org/10.1371/journal.pone.0196348.g020>

6.4 Performance of proposed noise removal methods for long duration tasks

To check the scalability aspects for long duration task on the proposed eye tracking noise removal methods, the NG task is modified. The inter-number spacing is selected to be 50 pixels (as this is the least spacing). The total duration of the task is set to 15 minutes approximately in which 9 random single digit numbers (4 odd numbers) appear on the screen at a time. The participants are expected to gaze and click on these 4 odd numbers only. After 4 clicks a new set of numbers appears on the screen. Totally 3 participants were taken for this case. Fig 20 shows the variable error related parameters-smoothness and closeness for one participant P1.

The participant P1 performed the task for a duration of 15 minutes approximately and hence, the results are shown by computing the parameters over a window length of 60 seconds. It is noted from Fig 20 that our variable error removal technique is able to reduce the noise (i.e. extracting dense fixation chunk from raw eye tracker data), whereas, the parameters corresponding to the raw data degrade considerably over time as the participant felt exhausted and stopped the experiment abruptly at the 14th minute. This observation (i.e. how subjective fatigue and exhaustion affects the variable error) also emphasizes the fact reported in [15]. In order to establish the fatigue factor on variable error, we have carried out same analysis for another participant P2 who was well-acquainted with the data capture procedures as he had participated several times during the initial phases of our experiment. From the subjective feedback, it was clear that he did not feel exhausted during the study and Fig 21 also supports the fact. The nature of variable error (see Fig 21) is somewhat constant owing to the raw data, whilst the proposed GSP + KF method is successful in handling the variable noise in contrast to the state-of-the-art methods.

Figs 20 and 21 truly justify how the proposed method is robust enough in handling the variable error induced by fatigue, head movement and exhaustion etc. Table 4 provides the consolidated results for the variable error correction with respect to the closeness and smoothness parameters for the proposed GSP + KF technique. It can be seen that the variation in the parameters for long duration task is within the 10% range of the short duration task. Table 5 shows the accuracy of correcting the systematic error for the short and long duration task.

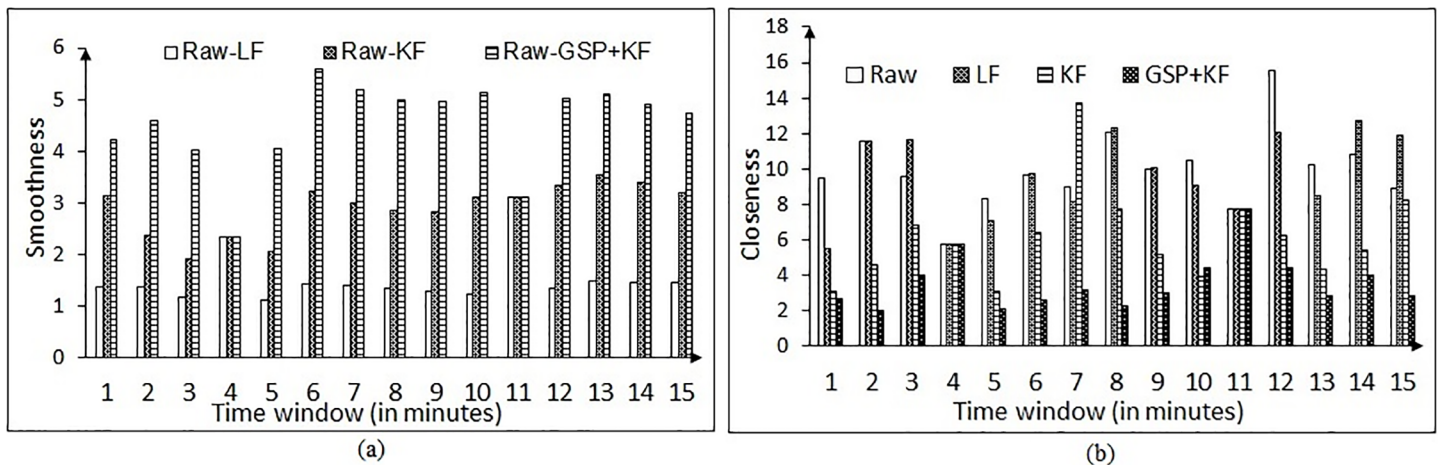


Fig 21. Removal of variable error for participant P2. (a) Smoothness parameter, (b) Closeness parameter.

<https://doi.org/10.1371/journal.pone.0196348.g021>

Table 4. Comparison of variable error correction in terms of closeness measure and smoothness ratio in short and long duration task using GSP + KF.

	Short Duration Task	Long Duration Task
Smoothness	3.79	3.96
Closeness	6.62	5.36

<https://doi.org/10.1371/journal.pone.0196348.t004>

Table 5. Comparison of average accuracy (%) in systematic error correction for short and long duration task.

	Short Duration Task	Long Duration Task
Path A	53.75	53.95
Path B	58.75	54.78
Path C	78.75	74.49
Path D	83.75	73.71

<https://doi.org/10.1371/journal.pone.0196348.t005>

Note that the accuracies are consistent bereft of the duration of the task while handling the systematic error. Path C and D perform better in denoising the systematic error.

7 Conclusions

The study aims at denoising a low-cost eye tracker in order to make it a perfect choice for the applications, such as rehabilitation, cognitive assessments, etc. The noise characteristics of a low resolution eye tracker are studied thoroughly and optimized approaches are designed to handle the errors associated with those errors. The algorithms are tested on 2 simple test stimuli and it is seen that our approach improves the overall performance of the system. In case of variable error, our proposed method reduces the dispersion of data points (i.e. closeness) by 48.98% and 59.53% in comparison with the raw data, for RR and the NG task, respectively. For systematic error removal, the results show improvements of about 17.86% and 15.25% over the raw data, for RR and NG tasks (taken average across all proposed paths for one time

calibration). The chosen test stimuli are closely related to the psychological tests and our results are motivating enough for the usage of eye tracker as a physiological sensor that can be used to extract more subject specific information such as working memory, attention or engagement, visual-motor coordination, etc., in real-time feedback for home-based applications. In addition to this, we have devised a one-time calibration protocol to avoid repeated calibration. Results confirm that the proposed approach gives satisfactory results in comparison to its multiple calibration scheme. Thus, it can be used successfully for patients who are unable to perform calibration due to some medical conditions. We have also evaluated our algorithms for long duration tasks and the results obtained are quite satisfactory. The system suits well for rehabilitation purpose. For the sake of rigorous scientific applications, the study needs to be further examined with detailed case studies governing the cognitive and behavioral aspects of eye movements research. In future we intend to increase the task duration further and study the effects on applications involving dynamic visual scenes like that of driving scenarios.

Acknowledgments

The authors like to thank all the participants for their cooperation and valuable time during the experiment.

Author Contributions

Conceptualization: Debatri Chatterjee, Kingshuk Chakravarty, Aniruddha Sinha.

Data curation: Rahul Dasharath Gavas, Sangheeta Roy, Soumya Ranjan Tripathy.

Formal analysis: Rahul Dasharath Gavas, Sangheeta Roy, Debatri Chatterjee, Soumya Ranjan Tripathy, Kingshuk Chakravarty.

Investigation: Debatri Chatterjee, Kingshuk Chakravarty, Aniruddha Sinha.

Methodology: Debatri Chatterjee, Kingshuk Chakravarty.

Project administration: Debatri Chatterjee, Aniruddha Sinha.

Supervision: Debatri Chatterjee, Aniruddha Sinha.

Writing – original draft: Rahul Dasharath Gavas, Sangheeta Roy, Debatri Chatterjee.

Writing – review & editing: Rahul Dasharath Gavas, Debatri Chatterjee, Kingshuk Chakravarty, Aniruddha Sinha.

References

1. Raudonis V, Simutis R, Narvydas G. Discrete eye tracking for medical applications. In: Applied Sciences in Biomedical and Communication Technologies, 2009. ISABEL 2009. 2nd International Symposium on. IEEE; 2009. p. 1–6.
2. Lin CS, Huan CC, Chan CN, Yeh MS, Chiu CC. Design of a computer game using an eye-tracking device for eye's activity rehabilitation. *Optics and lasers in engineering*. 2004; 42(1):91–108. [https://doi.org/10.1016/S0143-8166\(03\)00075-7](https://doi.org/10.1016/S0143-8166(03)00075-7)
3. Khasnobish A, Gavas R, Chatterjee D, Raj V, Naitam S. EyeAssist: A communication aid through gaze tracking for patients with neuro-motor disabilities. In: *Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2017 IEEE International Conference on. IEEE; 2017. p. 382–387.
4. Duchowski AT. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*. 2002; 34(4):455–470. <https://doi.org/10.3758/BF03195475>
5. Aslin RN. Infant eyes: A window on cognitive development. *Infancy*. 2012; 17(1):126–140. <https://doi.org/10.1111/j.1532-7078.2011.00097.x> PMID: 22267956

6. WHO. Global Health and Aging; 2017. Available from: http://www.who.int/ageing/publications/global_health.pdf
7. Giroud M, Jacquin A, Béjot Y. The worldwide landscape of stroke in the 21st century. *The Lancet*. 2014; 383(9913):195. [https://doi.org/10.1016/S0140-6736\(13\)62077-2](https://doi.org/10.1016/S0140-6736(13)62077-2)
8. Lagun D, Manzanares C, Zola SM, Buffalo EA, Agichtein E. Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *Journal of neuroscience methods*. 2011; 201(1):196–203. <https://doi.org/10.1016/j.jneumeth.2011.06.027> PMID: 21801750
9. Chan F, Armstrong IT, Pari G, Riopelle RJ, Munoz DP. Deficits in saccadic eye-movement control in Parkinson's disease. *Neuropsychologia*. 2005; 43(5):784–796. <https://doi.org/10.1016/j.neuropsychologia.2004.06.026> PMID: 15721191
10. Kasprowski P, Harezlak K. Using non-calibrated eye movement data to enhance human computer interfaces. In: *Intelligent Decision Technologies*. Springer; 2015. p. 347–356.
11. Nyström M, Andersson R, Holmqvist K, Van De Weijer J. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods*. 2013; 45(1):272–288. <https://doi.org/10.3758/s13428-012-0247-4> PMID: 22956394
12. Hornof AJ, Halverson T. Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers*. 2002; 34(4):592–604. <https://doi.org/10.3758/BF03195487>
13. Špakov O. Comparison of eye movement filters used in HCI. In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM; 2012. p. 281–284.
14. Zhang Y, Hornof AJ. Mode-of-disparities error correction of eye-tracking data. *Behavior research methods*. 2011; 43(3):834–842. <https://doi.org/10.3758/s13428-011-0073-0> PMID: 21487905
15. Ramdane-Cherif Z, Naït-AliNait-Ali A. An adaptive algorithm for eye-gaze-tracking-device calibration. vol. 57. *IEEE*; 2008. p. 716–723.
16. Blignaut P, Wium D. Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior research methods*. 2014; 46(1):67–80. <https://doi.org/10.3758/s13428-013-0343-0> PMID: 23609415
17. Li D, Babcock J, Parkhurst DJ. openEyes: a low-cost head-mounted eye-tracking solution. In: *Proceedings of the 2006 symposium on Eye tracking research & applications*. ACM; 2006. p. 95–100.
18. Hennessey C, Duchowski AT. An open source eye-gaze interface: Expanding the adoption of eye-gaze in everyday applications. In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ACM; 2010. p. 81–84.
19. Zugal S, Pinggera J. Low-Cost Eye-Trackers: Useful for Information Systems Research? In: *International Conference on Advanced Information Systems Engineering*. Springer; 2014. p. 159–170.
20. Gavas R, Roy S, Chatterjee D, Tripathy SR, Chakravarty K, Sinha A, et al. Affordable Sensor based Gaze tracking for Realistic Psychological Assessment. In: *Engineering in Medicine and Biology Society (EMBC), 2017 IEEE 39th Annual International Conference of the*, To be published. *IEEE*; 2017.
21. Bohme M, Meyer A, Martinetz T, Barth E. Remote eye tracking: State of the art and directions for future development. In: *Proc. of the 2006 Conference on Communication by Gaze Interaction (COGAIN)*; 2006. p. 12–17.
22. Barz M, Bulling A, Daiber F. Computational modelling and prediction of gaze estimation error for head-mounted eye trackers. *DFKI ResearchReports*. 2015; 1(1).
23. Duchowski A. *Eye tracking methodology: Theory and practice*. vol. 373. Springer Science & Business Media; 2007.
24. Deng LY, Hsu CL, Lin TC, Tuan JS, Chang SM. EOG-based Human-Computer Interface system development. *Expert Systems with Applications*. 2010; 37(4):3337–3343. <https://doi.org/10.1016/j.eswa.2009.10.017>
25. Stampe DM. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*. 1993; 25(2):137–142. <https://doi.org/10.3758/BF03204486>
26. Zhang Y, Gellersen H, Bulling A. *Eye tracking and gaze interface design for pervasive displays*. Lancaster University; 2015.
27. Noris B, Benmachiche K, Billard A. Calibration-free eye gaze direction detection with gaussian processes. In: *Proceedings of the International Conference on Computer Vision Theory and Applications*. *LASA-CONF-2007-018*; 2008.
28. Johansen SA, San Agustin J, Skovsgaard H, Hansen JP, Tall M. Low cost vs. high-end eye tracking for usability testing. In: *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM; 2011. p. 1177–1182.

29. Vidal M, Bulling A, Gellersen H. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. ACM; 2013. p. 439–448.
30. Fuhr W, Tonsen M, Bulling A, Kasneci E. Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art. *Machine Vision and Applications*. 2016; 27(8):1275–1288. <https://doi.org/10.1007/s00138-016-0776-4>
31. Olsson P. Real-time and offline filters for eye tracking; 2007.
32. van der Kamp J, Sundstedt V. Gaze and voice controlled drawing. In: Proceedings of the 1st conference on novel gaze-controlled applications. ACM; 2011. p. 9.
33. Veneri G, Federighi P, Rosini F, Federico A, Rufa A. Influences of data filtering on human–computer interaction by gaze-contingent display and eye-tracking applications. *Computers in Human Behavior*. 2010; 26(6):1555–1563. <https://doi.org/10.1016/j.chb.2010.05.030>
34. Gu J, Meng M, Cook A, Faulkner MG. Analysis of eye tracking movements using fir median hybrid filters. In: Proceedings of the 2000 symposium on Eye tracking research & applications. ACM; 2000. p. 65–69.
35. Sauter D, Martin B, Di Renzo N, Vomscheid C. Analysis of eye tracking movements using innovations generated by a Kalman filter. *Medical and biological Engineering and Computing*. 1991; 29(1):63–69. <https://doi.org/10.1007/BF02446297> PMID: 2016922
36. Abd-Almageed W, Fadali MS, Bebis G. A non-intrusive Kalman filter-based tracker for pursuit eye movement. In: American Control Conference, 2002. Proceedings of the 2002. vol. 2. IEEE; 2002. p. 1443–1447.
37. Grindinger T. Eye movement analysis & prediction with the kalman filter. Clemson University; 2006.
38. Komogortsev OV, Khan JI. Kalman filtering in the design of eye-gaze-guided computer interfaces. In: International Conference on Human-Computer Interaction. Springer; 2007. p. 679–689.
39. Vadillo MA, Street CN, Beesley T, Shanks DR. A simple algorithm for the offline recalibration of eye-tracking data through best-fitting linear transformation. *Behavior research methods*. 2015; 47(4):1365–1376. <https://doi.org/10.3758/s13428-014-0544-1> PMID: 25552423
40. Bignaut P. Using smooth pursuit calibration for difficult-to-calibrate participants. *Journal of Eye Movement Research*. 2017; 10(4).
41. Santini T, Fuhr W, Kasneci E. Calibme: Fast and unsupervised eye tracker calibration for gaze-based pervasive human-computer interaction. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM; 2017. p. 2594–2605.
42. Pygame. Pygame; 2017. Available from: <http://www.pygame.org/lofi.html>
43. Brand A, Jolles J, Gispen-de Wied C. Recall and recognition memory deficits in depression. *Journal of affective disorders*. 1992; 25(1):77–86. [https://doi.org/10.1016/0165-0327\(92\)90095-N](https://doi.org/10.1016/0165-0327(92)90095-N) PMID: 1624647
44. Amaresha AC, Danivas V, Shivakumar V, Agarwal SM, Kalmady SV, Narayanaswamy JC, et al. Clinical correlates of parametric digit-symbol substitution test in schizophrenia. *Asian journal of psychiatry*. 2014; 10:45–50. <https://doi.org/10.1016/j.ajp.2014.03.010> PMID: 25042951
45. Rao SL, Subbakrishna D, Gopukumar K. NIMHANS neuropsychology battery-2004, manual. National Institute of Mental Health and Neurosciences; 2004.
46. Rhodes MG, Castel AD. Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of experimental psychology: General*. 2008; 137(4):615. <https://doi.org/10.1037/a0013684>
47. Calvo A, Chiò A, Castellina E, Corno F, Farinetti L, Ghiglione P, et al. Eye tracking impact on quality-of-life of ALS patients. In: International Conference on Computers for Handicapped Persons. Springer; 2008. p. 70–77.
48. Riby DM, Doherty MJ. Tracking eye movements proves informative for the study of gaze direction detection in autism. *Research in Autism Spectrum Disorders*. 2009; 3(3):723–733. <https://doi.org/10.1016/j.rasd.2009.02.001>
49. Elahipanah A, Christensen BK, Reingold EM. What can eye movements tell us about Symbol Digit substitution by patients with schizophrenia? *Schizophrenia research*. 2011; 127(1):137–143. <https://doi.org/10.1016/j.schres.2010.11.018> PMID: 21147521
50. Salvucci DD, Goldberg JH. Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the 2000 symposium on Eye tracking research & applications. ACM; 2000. p. 71–78.
51. Tafaj E, Kasneci G, Rosenstiel W, Bogdan M. Bayesian online clustering of eye movement data. In: Proceedings of the symposium on eye tracking research and applications. ACM; 2012. p. 285–288.

52. Kasneci E, Kasneci G, Kübler TC, Rosenstiel W. The applicability of probabilistic methods to the online recognition of fixations and saccades in dynamic scenes. In: Proceedings of the Symposium on Eye Tracking Research and Applications. ACM; 2014. p. 323–326.
53. Sen T, Megaw T. The effects of task variables and prolonged performance on saccadic eye movement parameters. *Advances in Psychology*. 1984; 22:103–111. [https://doi.org/10.1016/S0166-4115\(08\)61824-5](https://doi.org/10.1016/S0166-4115(08)61824-5)
54. Shuman DI, Narang SK, Frossard P, Ortega A, Vandergheynst P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*. 2013; 30(3):83–98. <https://doi.org/10.1109/MSP.2012.2235192>
55. Obolensky N. Kalman filtering methods for moving vehicle tracking. University of Florida; 2002.
56. Mills TC. Time series techniques for economists. Cambridge University Press; 1991.
57. Faragher R. Understanding the basis of the kalman filter via a simple and intuitive derivation [lecture notes]. *IEEE Signal processing magazine*. 2012; 29(5):128–132. <https://doi.org/10.1109/MSP.2012.2203621>
58. Friedman JH, Bentley JL, Finkel RA. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*. 1977; 3(3):209–226. <https://doi.org/10.1145/355744.355745>
59. EyeTribe. The Eye Tribe; 2017. Available from: <http://theeyetribe.com/theeyetribe.com/about/index.html>