USee: Ultrasound-based Device-free Eye Movement Sensing

Wen Cheng[†], Mingzhi Pang[†], Haoran Wan[‡], Shichen Dong[†], Dongxu Liu[†], Wei Wang[†]
[†] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[‡] Department of Computer Science, Princeton University, Princeton, USA
{wcheng, mzpang, scdong, dongxuliu}@smail.nju.edu.cn, haoran.w@princeton.edu, ww@nju.edu.cn

***Abstract*—Eye movements play a significant role in human-computer interaction and are widely recognized as an essential health indicator, making their detection both appealing and technically challenging. In this paper, we present a system named USee that achieves high-precision capture of weak and aperiodic eye movements by utilizing fine-grained and ubiquitous ultrasound signals, capturing both blinking and more subtle saccades. We first identify signal changes associated with eye movements by capturing the unique impact of blinking. Further, we establish a pioneering relationship between the residuals from signal decomposition and subtle eye movements. Utilizing innovative signal processing architectures, we mitigate interference and effectively extract eye movement features. Subsequently, we employ one-dimensional convolutional operations in place of signal cross-correlation, designing filters for motion category identification and a lightweight convolutional neural network for saccade direction classification. This enables our system to serve as a foundational sensing layer for eye movement tracking, applicable across diverse applications. We implement USee on both a research-purpose platform and a commodity Raspberry Pi. Extensive experimental results demonstrate the effectiveness of our system, achieving 91% accuracy in saccade recognition and 94% in blink detection. The system proves robust, even in challenging scenarios with strong interference, such as the presence of moving pedestrians.**

***Index Terms*—Wireless Sensing, Eye Movement Detecting**

I. Introduction

"The soul, fortunately, has an interpreter – in the eye." – Charlotte Bronte

Often poetically referred to as the windows to the soul, the eyes provide one of the most essential human senses – vision. The study of eye movements has long been a prominent topic for both academia and industry because high-fidelity eye movement estimation is vitally important in many scenarios, from assisting medical diagnosis to improving interactive experience. In the aspect of medical assistance, saccades, which are the rapid movements that sharply change the point of view and the most frequently occurring eye movements in humans, can serve as important indicators for diagnosing Parkinson's disease [1] and Alzheimer's disease [2]. Capturing eye movements, especially saccades, is crucial for human-computer interaction, enabling applications like user intention/status monitoring and Virtual Reality (VR) and Mixed Reality (MR) interaction. Eye movements reveal user attention, aiding in driver-state monitoring and productivity through saccade detection and reading speed estimation, as we discuss in Sec. IV-E. In VR/MR, eye movement is a key interaction method, already used in Commercial-Off-The-Shelf (COTS) products like Apple Vision Pro and Meta Quest Pro.

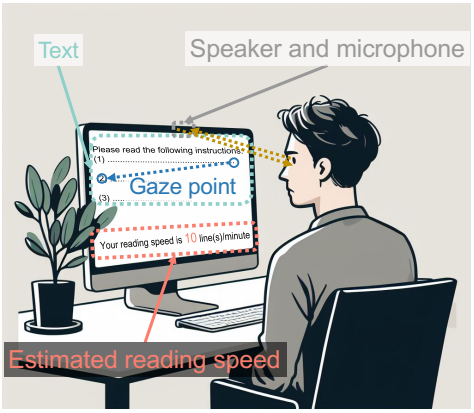


Fig. 1: USee's daily usage – reading speed estimation.

Existing methods for capturing eye movements can be roughly divided into three classes, specially-designed sensors, device-based and device-free methods, where each has its own drawbacks. Specially designed sensors, such as Electrooculography (EOG) and infrared sensors, can accurately detect a user's eye movement. However, this specially designed wearable solution incurs high costs and is not comfortable enough due to the extra power supply requirement [3]. In recent years, wireless sensing has emerged as a prominent field, distinguished by its lightweight device requirements and efficient recognition capabilities. Yet, research focusing on capturing eye movements using wireless sensing remains limited. For device-based sensing method, Smartlens [4] attaches dedicated antenna circuits to the contact lens, as well as the RF tags, and achieves eye movement direction recognition. However, this contact lens-based approach still requires additional devices and does not apply to individuals who do not use contact lenses. For device-free sensing methods, BlinkListener [5] and TwinkleTwinkle [6] employ ultrasonic signals in a contact-free manner to recognize blinks, which is coarse-grained and don't fully reflect the user's eye movement status, limiting their practicality. Therefore, it naturally leads to the question: Is there a method that *combines the advantages of being non-invasive and device-free while pushing the limits to achieve accurate fine-grained eye movement detection?*

In this work, we present USee, a system that leverages ubiq-

TABLE I: Comparison of blinking and saccade.

Distance (cm)	30	40	50	60
${R_{amp}}^2$	2.47	3.03	2.19	2.30
${Cor_{saccade}}^2$	0.35	0.56	0.65	0.87
${Cor_{blink}}^2$	0.11	0.14	0.21	0.30
SNR of blinks (dB)	17.63	13.14	7.52	5.4
SNR of saccades (dB)	7.58	5.35	3.33	1.66

uitous short-wavelength acoustic signals to accurately detect fine-grained saccade movements. This non-invasive solution can be seamlessly integrated into interactive devices (Fig. 1). To achieve finer-grained, device-free saccade estimation, we analyze the subtle changes induced by saccades and blinks using our intricately designed, innovative signal processing pipeline. Earlier methodologies [5], [6], which depended on absolute amplitude or threshold-based methods, proved inadequate for detecting substantially smaller saccades. Moreover, we reveal the relationship between signal decomposition residuals and subtle movements, enabling the detection of extremely delicate saccades. The direction of saccades is then further classified by a light-weight neural network. We summarize the following challenges USEE must address to achieve high-precision device-free eye movement detection:

- *Locating the signal corresponding to the eye movements.* The received signal is a mix of reflections from various objects at different distances, including interference from mouth and head movements, along with the eyes. To detect saccade patterns, we must determine the relative position of the eyes to the receiver so that we can identify signal changes resulting from eye movements in the received signal. We leverage the characteristic that saccades and blinks cause distinct signal changes to address this challenge, as detailed in Sec. III-D1.
- *Extracting the extremely fragile and aperiodic eye movements.* The saccade movement is considerably weaker than a blink, with a tiny displacement of 1-2 mm [7]. From our real-world experiments (Table I)[1], we observe that the signal pattern caused by saccades is significantly more subtle compared to that of blinks, as measured by Signal-to-Noise Ratio (SNR) and three other metrics[2]. Additionally, the signal changes induced by saccades are less distinctive and more easily obscured by background noise. To address this challenge, we introduce an innovative approach that reveals the relationship between signal decomposition residuals and subtle eye movements, providing a reliable indicator for detecting saccades and blinks, as detailed in Sec. III-D3.
- *Distinguishing between saccades and blinks.* After extracting signals related to eye movements, we obtain residuals that include both saccades and blinks. However, the limited information within these residuals makes it challenging to differentiate between the two actions. Commencing from the distinct nature of saccades and blinks, we aim to design a filter to differentiate between them. We observe that one-dimensional convolution operations incorporate a correlation operator, aligning well with our approach. Thus, we achieve this goal through a one-dimensional convolutional neural network (CNN), as detailed in Sec. III-D4.

In summary, the main contributions of our work are as follows.

(1) To the best of our knowledge, USEE is the first device-free deployed system employing acoustic signals for fine-grained eye movement capturing and saccade direction classification with high accuracy and reliability.
(2) We present a comprehensive signal processing pipeline for extracting eye movement features. Specifically, we uncover the relationship between the residuals from the signal decomposition method (as opposed to the decomposition results themselves) and subtle eye movements, enabling effective extraction of the signals associated with eye movements.
(3) We implement the USEE on both a research-focused hardware platform as well as on a widely used consumer device, the Raspberry Pi. Extensive experiments validate its reliability and accuracy.

II. Related Work

A. Traditional Eye Movement Detection Methods

1) EOG sensor-based: EOG sensors measure biopotential signals induced by changes in the cornea-retina dipole characteristics during eye movement. These sensors have been widely explored for eye blink detection and eye movement tracking over decades [8], [9]. Existing EOG-based solutions usually use glasses as the sensor carrier, where EOG sensors and PCB are attached to the custom material frame [3]. Commercial EOG glasses, such as JINS MEME glasses [10], alleviate the discomfort of capturing signals to some extent. However, the customized hardware comes with an additional price tag and has to cope with energy-hungry computations. Conversely, acoustic or camera-based solutions can be deployed on commercial devices without additional hardware.

2) Infrared-based: Nowadays, infrared sensors have been integrated within commercial glasses or headsets for eye tracking, as in Apple Vision Pro [11]. Despite the advantages of infrared light sensors, such as smaller size and lower energy consumption, the potential safety hazard of exposing the eye to infrared light remains a concern [12], [13]. The comfort of wearing the head-mounted devices is also an important factor, and prolonged wearing tends to cause cervical spine fatigue [14]. Our approach eliminates the need to wear any additional equipment, and since the volume of the sound waves we employ is well below safe exposure limits [15], there are no safety concerns even for protracted usage.

3) Camera-based: Camera-based solutions have been widely applied in eye movement detection systems [16].

[1]The experiment setup is similar as in Sec. IV-A, data was collected at 10 cm intervals within a distance range of 30$\sim$60 cm.

[2]The ratio of signal amplitude changes induced by blinks and saccades (R_{amp}). The correlation coefficients, Cor_{blink} and $Cor_{saccade}$, are the correlation between the power spectra of eye movements and those of noise to measure the relationship between eye movements and noise.

Typical capturing procedure involves first recognizing face landmarks from the image, cropping image segments containing the eyes, and then tracking eye movements using back-end algorithms [17]. Such a processing pipeline relies on a clear and unobstructed view of the face [17], therefore the tracking accuracy of the existing vision-based methods decreases in the case of wearing a mask [18]. Moreover, vision-based methods tend to be computationally intensive and have privacy concerns [19]. In contrast, our scheme does not require the face to be completely unobstructed, nor does it have any requirement for ambient light.

B. Acoustic-based Eye Movement Detection Methods

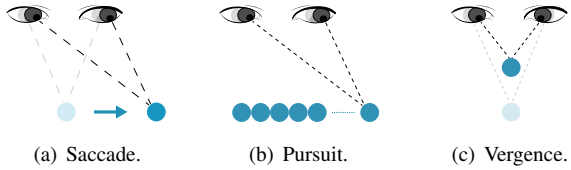
Several recent works focus on sensing blinks using acoustic signals. BlinkListener [5] utilizes FMCW ultrasonic signals for blink detection, employing an innovative approach based on frequency bin processing. It observes signal amplitude variations caused by changes in reflective materials and proposes an optimal viewing angle to maximize the amplitude changes induced by blinks, effectively distinguishing them from background interference. Another innovative work, TwinkleTwinkle [6], constructs an HCI system using FMCW signals to interact with the blinking pattern. It mitigates interference through phase difference and distinguishes various user blink habits using adaptive thresholds. Combining a vote-based method, it maps the blink pattern to symbols like ASCII for interaction. However, our research deviates from the above work by concentrating on the detection of saccades, eye movements that are more subtle than blinking. GazeTrak [20] employs four microphones and one speaker attached to each side of the glasses for high-precision and high-sensitivity monitoring of eye movements. In contrast, our study focuses on device-free eye movement tracking, eliminating the need for wearable devices. We achieve precise eye movement monitoring using only one microphone and one omnidirectional speaker.

III. Acoustic Sensing System for Eye Movements

In this section, we first introduce the typical forms of eye movements, with saccades being the most frequent eye movement. We then provide a detailed description of the design of each component in USee.

A. Background

Typically, three types of voluntary eye movements are considered primary: saccades, pursuit movements, and vergence movements. Saccades are rapid conjugate eye movements where eyes moves from the center to the left or the right (Fig. 2(a)). Pursuit movements are much slower, smoother conjugate tracking movements of the eyes (Fig. 2(b)). Vergence movements are disconjugate, where two eyes move in opposite directions (Fig. 2(c)). They typically occur when tracking a target from a distance to close range or vice versa. Of the three forms of movement above, saccades are the *most frequent* eye movement in daily life [21]. In addition to those subtle eye motions, blinking is a much more drastic action of the eyelid and eyeball muscles to clean and refresh our eyes. Therefore, given the frequent occurrence of saccades and the significance of blinks, USee primarily captures saccades and blinks.



(a) Saccade. (b) Pursuit. (c) Vergence.

Fig. 2: Forms of eye movements.

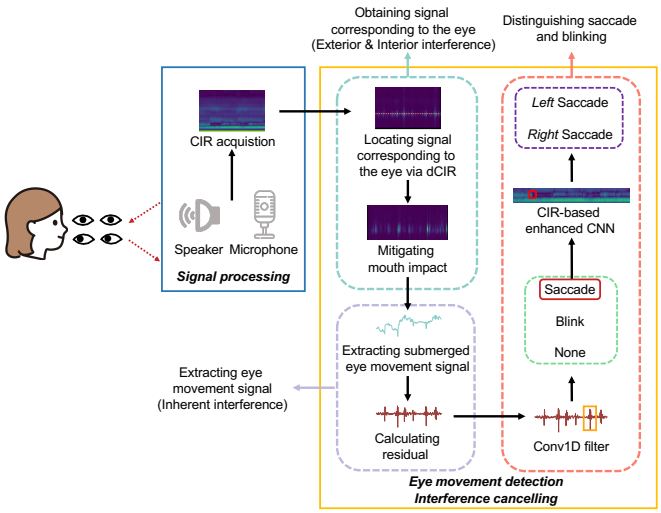
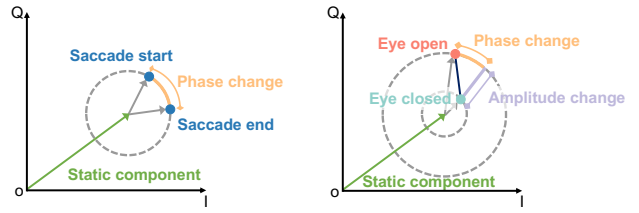


Fig. 3: Overview design of the USee.

B. System Overview

Our system comprises two main components: the signal processing module and the eye movement detection module, as shown in Fig. 3. The signal processing module handles ultrasound signal modulation, demodulation, and Channel Impulse Response (CIR) computation. The eye movement detection module mitigates the interference in the raw CIR, extracts the eye movement patterns and classifies the eye movement types. For the interference in the raw CIR, USee addresses three types of interference: exterior, interior, and inherent (please refer to Sec. III-D and Fig. 5 for definition). We first use blinking's unique impact on signal amplitude to identify the reflective distance related to eye movements, mitigating interference from mouth movements to eliminate exterior and interior interferences. After isolating eye movement-related reflections, we apply Variational Mode Decomposition (VMD) to minimize inherent interference. For the movement types classification, recognizing the non-periodic, impulse-like nature of eye movement reflections as beneficial, we leverage the correlation between signal decomposition residuals and subtle eye movements, extracting features from these residuals, as the classification features. These features are then input into a one-dimensional convolutional filter for motion classification. To further identify the saccades directions, we design a convolutional neural network based on CIR to enhance saccade recognition, differentiating between left and right saccades.



(a) Phase change caused by saccade. (b) Amplitude and phase changes caused by blink.

Fig. 4: Viewing eye movements in the I-Q vector space.

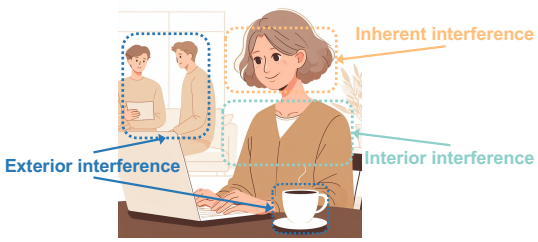


Fig. 5: Different interferences when capturing eye movement.

C. Signal Processing Module

1) Signal Modulation: To balance signal energy and achieve path separation, USEE transmits periodic high-energy signals and leverages the autocorrelation properties of ZC sequences for modulation in the frequency domain. The baseband signal is shifted to the desired frequency using a high-frequency carrier, creating a real modulated signal through conjugate symmetry. The final time-domain signal is obtained via IFFT.

2) Signal Demodulation and CIR Acquisition: Upon receiving the signal, the microphone segments it into frames, performs FFT and matches the baseband frequency components. The conjugate of the modulation sequence is multiplied by the received signal in the frequency domain, and IFFT is applied to acquire the time-domain CIR [22], [23].

D. Eye Movement Detection Module

We categorize interferences encountered during eye movement recognition into three types: exterior, interior, and inherent, as shown in Fig. 5. Our eye movement detection module incorporates methods to address each type.

1) Exterior interference: Exterior interference refers to reflections from surrounding objects, both static (like tabletop ornaments) and moving (like passing individuals).

Solution. CIR measurements provide information on propagation path length and reflected signal strength, forming a 2D CIR map like Fig. 6. The horizontal axis in the map denotes time ranges, while the vertical axis denotes distance ranges. The time-domain resolution of 0.025 s in the CIR map ensures a 40 Hz sampling rate, which is sufficient for capturing the subtle movement of the eye[3]. In frequency domain, acoustic signal with a 6 kHz bandwidth can distinguish objects over 5.7 cm apart. The 2D CIR clearly shows how the signal

[3]Table. III provides the detailed parameters of the signals.

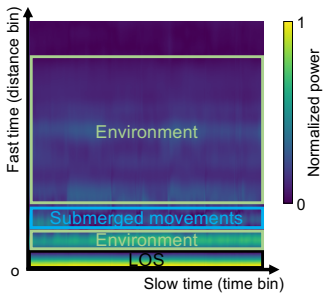


Fig. 6: CIR example for eye tracking.

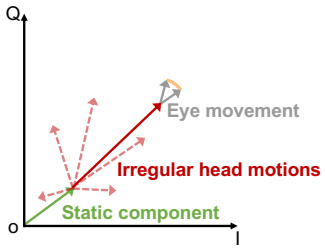


Fig. 7: Head motions influences.

changes with time at different distances, which makes it simple to remove exterior interference by filtering long-distance reflections.

2) Interior interference: After eliminating exterior interference, we focus on locating eye-related signals and mitigating potential interior interference. Interior interference refers to reflections from other regions within the human body, excluding external factors. These can originate from the torso, such as thoracic cavity fluctuations during respiration, or from nearby facial features, like mouth movements during speech.

Solution. The CIR contains signal variations related to eye movements, and locating and isolating these signal patterns is vital for detection yet challenging. Fig. 4 shows how blinking manifests as changes in signal amplitude and phase in the I-Q vector space. Compared to the saccade, which is caused by corneal movement and has a relatively small change in path, a blink not only causes a change in the distance of the signal propagation path, but also causes a change in the reflective material, resulting in a more significant change in both the signal's amplitude and phase. Considering that blinks and saccades occur in the same location and are mutually exclusive as the eyes can perform only one action at a time. Therefore, the unique characteristic of blinking could serve as a crucial indicator for locating signals related to eye movements in the 2D CIR. To highlight signal changes from blinking, we use differentiated CIR (dCIR) to remove static components. This allows detection based on dCIR's absolute amplitude, helping to determine the relative distance between the eyes and the receiver and capture key eye movement signal patterns. The 2D CIR's distance information allows for eliminating interference from distant torso activities. However, interference from nearby sources, like mouth movements, remaining unaddressed. The distance between the human mouth and eyes is approximately 6.5 cm [24], which is close to the limit of ultrasound signal resolution. As a result, current methods often passively remove interference from mouth movements by discarding potentially corrupted frames [5]. This approach undoubtedly results in the loss of information related to eye movements in the signal.

We design a multi-bin re-combination approach to mitigate the interference of mouth movements on eye movement recognition based on our observation of eye movement patterns. Denote the signal pattern induced by eye movements as s_e

TABLE II: Frequencies and displacements for different movements.

Motions	Intervals / Frequency	Displacement
Eye movements (Saccade)	Aperiodic, interval: 0.3–2 s [25]	1–2 mm
Eye movements (Blinking)	Aperiodic, interval: 3–4 s [26]	Millimeter level
Head motions (BCG)	1–1.5 Hz [27]	Millimeter level
Head motions (Torso-induced)	0.2–2 Hz [28]	Subcentimeter level
Head motions (Natural status)	0.4–3.5 Hz [29]	Centimeter level

and the signal pattern induced by mouth movements as s_m. At the bin corresponding to the distance of the eyes, the combined signal is approximately the weighted sum of the two patterns, $S_{eye} = w_e \times s_e + w_m \times s_m$, where w_e and w_m are the weight for eye and mouth movement patterns. Similarly, at the bin corresponding to the distance of the mouth, the combined signal pattern can be expressed as $S_{mouth} = w'_e \times s_e + w'_m \times s_m$. Usually, the reflection signal strength of the mouth dominates the inference from eye movements at the bin corresponding to the distance of the mouth, given the close distance and sizes. Hence by assuming that $w'_m \gg w'_e$, we have $S_{mouth} \approx w'_m \times s_m$. Therefore, we use the least squares method to extract the eye movement by minimizing the impact of mouth movements, *i*.e., estimating a parameter θ that satisfies $\arg\min_\theta (S_{eye} - \theta \times S_{mouth})$.

However, due to the aperiodical nature of eye movements and mouth movements, we need to perform such estimation on short time windows to compensate for head movements that may change the distance between the mouth and the eyes. We chose a window size of 20 frames (0.5 s in time) as the length of the dCIR processing based on experimental results.

3) Inherent interference: Eye movements generate tiny reflected signal patterns that are fragile and highly sensitive to interference from other movements. Even minor motions can significantly reduce the effectiveness of eye movement capture. Involuntary movements, such as natural head swaying and torso-driven displacements during respiration, persist throughout the eye-tracking process. We refer to these as inherent interference. The primary source of inherent interference is the unconscious displacement of the head, driven by torso motion during respiration, with millimeter-level displacements at the respiratory rate. Additionally, the pulsatile effects of the heart, known as Ballistocardiography (BCG) [30], cause similar head displacements during each cardiac cycle. Natural micro-movements of the head during eye movement also contribute, with sub-centimeter displacements at frequencies below 3 Hz [31]. Table II summarizes the displacement amplitudes and motion frequencies for both inherent interference and eye movements.

Solution. Eye movements are extremely subtle compared to some inherent interference, *e.g.* natural head movements causing displacements that are orders of magnitude larger. This results in a mixed representation of eye movement signal in the I-Q vector space, as shown in Fig. 7. Moreover, the overlapping frequencies of the four signal components make it nearly impossible to isolate them using a simple band-pass filter. Therefore, a more sophisticated method is required to separate these signal components effectively.

The aperiodic nature of eye movements inherently limits the ability to extract meaningful information directly from the frequency domain. Signal decomposition methods aim to separate mixed signals into components with different bandwidths. Traditional methods, such as EEMD [32] and CEEMDAN [33], extract signal components (IMFs) from high to low frequencies through a sifting process based on the original signal. However, they often group eye movement signals with high-frequency noise into the same component, making it difficult to effectively isolate eye movement information from the IMFs. Recent methods like VMD [34] address this by optimizing the reconstruction of the original signal from components while imposing constraints on center frequency and bandwidth. Despite these advancements, these methods still struggle to extract high-SNR information related to eye movements from the decomposition results. The limitations are illustrated on the left of Fig. 8, which shows the components obtained through the three decomposition methods mentioned. Since the highest frequency component most effectively captures eye movement information, we focus on illustrating this specific component. The decomposition results show a mixture of signal fluctuations from both eye movements and noise, leading to indistinct features and time misalignment, which complicates the overall interpretation. However, this does not mean that extracting signals related to eye movements is impossible. Let us take a different perspective. Given that eye movements are brief and resemble impulse signals, their frequency components span the entire frequency domain, which precludes the possibility of complete decomposition. We now define the residual of signal decomposition as follows to examine this non-decomposable portion:

$$R(t) = f(t) - \sum_{i}^{K} IMF_i, \tag{1}$$

where K represents the maximum value of the subscript for IMFs. Eq. (1) implies that the residual is defined as the result obtained by subtracting all IMF components from the original signal. This component precisely represents the portion of the signal that cannot be decomposed, corresponding to the high-frequency and high-energy components of the original signal [34], aligning with the frequency-domain and energy characteristics of eye movements. The right of Fig. 8 illustrates the residuals obtained from three signal decomposition methods. Notably, in the VMD residual, signals related to eye movements stand out distinctly from background noise. The characteristics of eye movements are clearly highlighted, demonstrating a clear separation between these signals and the surrounding noise. This is due to the computational characteristics of the VMD method itself. In contrast, the other two methods, which aim to fully recover the original signal, mix eye movement signals into high-frequency components, rendering the residuals ineffective in accurately depicting eye movement signals. Thus, we establish a connection between signal decomposition residuals and subtle eye movements,

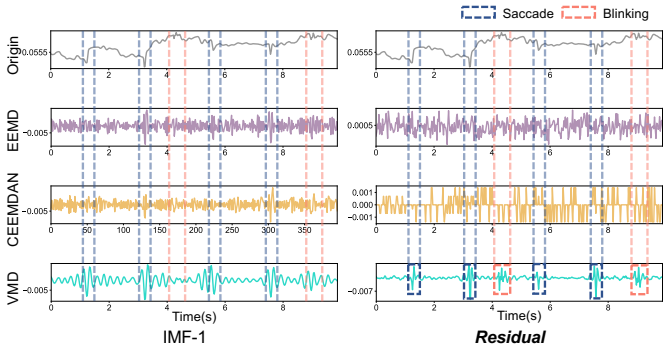


Fig. 8: The IMF-1 and residual of different signal decomposing methods.

successfully extracting eye movement characteristics from the underlying signal.

4) Distinguishing saccade and blinking: Although the VMD residual effectively reduces interference and highlights eye movement signals, distinguishing between saccades and blinking remains a challenge. We attempt to use linear classification methods based on thresholds, a common approach in other systems. However, experimental results show limited effectiveness in differentiating these two actions. The accuracy of CFAR for saccades reached only 44%, while for blinking it was 53%. Despite this, we observe differences in action duration and energy intensity in the reflection patterns of saccades and blinking. Our approach involves training a series of filters to make them responsive to one action while rejecting the other, a computation process resembling signal cross-correlation. The one-dimensional convolution operation (*conv1D*) in neural networks closely resembles the aforementioned cross-correlation computational process. When provided with a signal of length L and C_{in} channels, the i^{th} output channel of one-dimensional convolution is calculated as: $out_i = \sum_{j=1}^{C_{in}} weight_{i,j} * in_j$, here, $*$ denotes the cross-correlation operator. The shape of the outcome is (C_{out}, L_{out}), where $L_{out} = L_{in} - window\_size + 1$. We determine the window size based on the duration of eye movements and the signal's time resolution. By incorporating cross-correlation within the one-dimensional convolution, filter design is effectively transformed into training a simple neural network model. We use the cross-entropy function as the loss function to train the network. Detailed model parameters and experimental results are discussed in Sec. IV.

5) Saccade direction classification: After obtaining saccades, we aim to refine the recognition granularity further. Compared to CIR, the information contained in the residuals is insufficient for further differentiation. Therefore, we use CIR as the input data and design a convolutional neural network model, widely used in the field of ultrasound action classification [35], to classify saccades into left and right directions. Specifically, upon detecting a saccade, we segment the CIR of the single signal related to the eyes by selecting the preceding and following 10 distance bins within the action period. This results in a two-dimensional matrix data with the shape of $20 \times window\_size$. We then input this matrix into a lightweight neural network model with two two-dimensional convolutional layers for direction classification. Experimental results confirm that this approach achieves a high level of accuracy. Detailed experimental results are provided in Sec. IV.

TABLE III: Parameters of sending signals on two platforms.

Parameters	NI USB-6356	Raspberry Pi 4B
Sample frequency	96 kHz	48 kHz
Center frequency	38 kHz	19 kHz
Frame length	2400	1200
Bandwidth	6 kHz	6 kHz

IV. Evaluation

A. Implementation

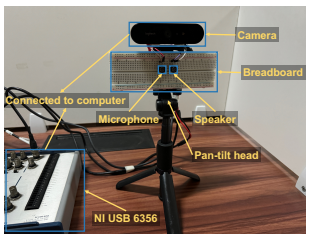
We implement USee on both the research-oriented NI USB-6356 and the consumer-grade Raspberry Pi 4B [36], as shown in Fig. 9. Both platforms use a single-transmitter, single-receiver setup, with the microphone placed approximately 2–3 cm away from the speaker. For the Raspberry Pi, the Respeaker 4-mic linear array is used, but only one channel serves as the input.

The NI platform is used as the benchmark for its flexibility in exploring experimental configurations, allowing us to investigate various potential applications of USee, such as integration in eyeglasses, and demonstrate deployment possibilities in embedded wearable devices like VR headsets. Table. III provides the detailed parameters of the signals.

The signal processing is performed on a laptop with an AMD Ryzen 5600H CPU and 16GB of RAM using MATLAB. The digital filter and saccade direction classification model training is developed using Python and PyTorch. In the neural network filter model, the kernel size (window size) of the 1D convolution layer is set to 19, which can cover the temporal span of both saccade and blink actions. In the enhanced saccade direction classification model, the number of 2D convolution layers is set to 2, and the kernel size for each is 3. For the NI platform, the training dataset comprises approximately 1,600 instances of blinks and 1,100 instances of saccades, while the Raspberry Pi platform's independent training dataset includes about 1,300 blinks and 1,000 saccades. For the overall performance evaluation, the ratio of the training set, the validate set, and the test set is 8:1:1. The test dataset used in the subsequent impact factor experiments is independent and not included in the above data set. Ground truth of eye movement is obtained through manual labeling using video streams captured by a camera fixed at the top of the tripod.

B. Experiment Setup

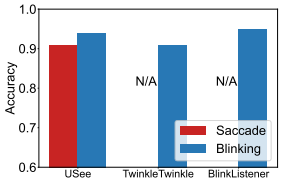
We recruited eight participants when collecting the evaluation data set, each with diverse backgrounds and eye movement habits. Each sample of the data set contains 10 seconds of eye movement signals. Participants independently decided



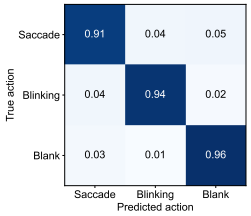
(a) The NI sensing platform.

(b) The Raspberry Pi platform.

Fig. 9: Experimental setup and environment.



(a) Performance comparison with counterparts.



(b) Confusion matrix of the detection.

Fig. 10: The overall performance of USEE.

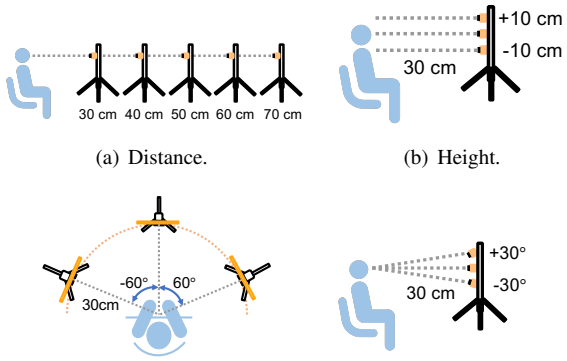
on the eye movements in each data sample, typically including 3−6 saccades with 1−3 blinks. Our data set is collected in normal lab environments, with tables and chairs placed around the test site. Unless otherwise specified, the sensing unit is placed at the same height as the participants' eyes as shown in Fig. 9(b). For the mouth mitigation experiment, participants are required to count from 1 to 100 at their usual conversational speed and volume, with the sound level typically maintained within the range of 60∼70 dB.

C. Overall Performance

We employ a detailed accuracy measure by determining if each temporal frame correctly detects eye movements. On the NI platform, USEE achieves 91% accuracy for saccade recognition and 94% for blink recognition. The confusion matrix in Fig. 10(b) shows that errors mainly occur when saccades and blinks are confused. When considering only the presence of eye movement, the accuracy rises to 96%. On the Raspberry Pi platform, USEE achieves 90% accuracy for saccades and 93% for blinks, proving its feasibility on consumer devices. Additionally, the system achieves 92% accuracy for classifying saccade directions (left and right).

We benchmark USEE against BlinkListener and TwinkleTwinkle, two leading ultrasound-based blink detection systems. USEE matches BlinkListener's blink recognition accuracy on both platforms and outperforms TwinkleTwinkle. Notably, USEE also accurately detects saccades, a capability not present in the other systems.

In the presence of sustained mouth movements, the unprocessed data results in low recognition accuracies, with only 28% for saccades and 36% for blinks after processing. However, after applying our interference mitigation method and subsequent processing, the recognition accuracies significantly improved to 41% for saccades and 62% for blinks, demonstrating the effectiveness of our proposed approach.



(a) Distance. (b) Height.

(c) Angle. (d) Orientation.

Fig. 11: Different relative positions of the sensing setup.

D. Impact Factors

We assess USEE's performance under varying relative positions, as shown in Fig. 11. Distances, heights, angles, and orientations are adjusted at consistent intervals, with the eye-facing direction set to 0 in the directional symmetry experiment.

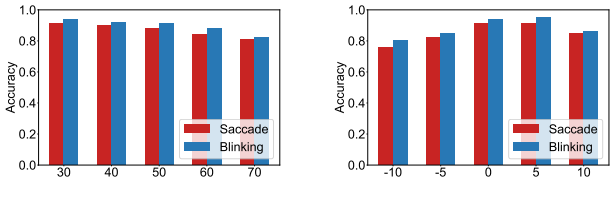
1) Distance: As shown in Fig. 12(a), our system achieves 91% accuracy for saccades and 94% for blinks at 30 cm. Even at 70 cm, accuracy remains around 80%, which aligns with recommended interaction distances for electronic devices [37]. On a Raspberry Pi, sensing accuracy reaches 86% for saccades and 88% for blinks. The lower accuracy is due to reduced SNR as distance increases. Future research with directional or high-power transmitters could further enhance accuracy.

2) Height: As shown in Fig. 12(b), raising the device by 5 cm improves accuracy to 91% for saccades and 95% for blinks by reducing interference from lower body reflections. However, placing the device too high weakens the signals from eye movements, reducing recognition performance. Lowering the device by 10 cm increases interference, decreasing accuracy to 76% for saccades and 80% for blinks.

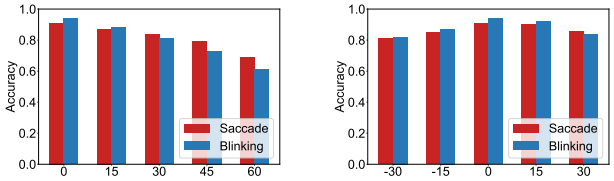
3) Angle: We can observe from Fig. 12(c) that the system maintains accuracy within an acceptable range even when the bias is within 30 degrees, reaching 84% accuracy for saccades and 81% accuracy for blinks. As the bias angle increases, performance further deteriorates, mainly due to the reduction of eye movement information in the received signal.

4) Orientation: As shown in Fig. 12(d), adjusting the device orientation upward has a minimal impact on performance. A 15-degree increase achieves 90% accuracy for saccades and 92% for blinks, while a 15-degree decrease has a similar effect as a 30-degree upward adjustment. Lowering the orientation by 30 degrees reduces accuracy to 81% for saccades and 82% for blinks due to increased interference from other facial regions.

5) Different Type of Exterior Interference: As shown in Fig. 13(a), we conducted three experiments to simulate real-world interference: (i) An interfering user walking 40 cm to

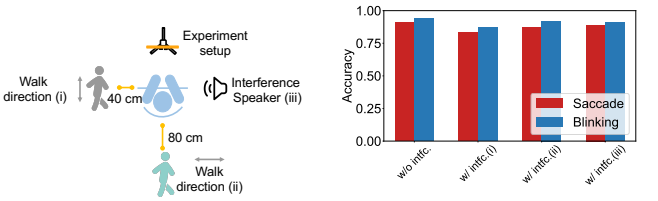


(a) Distance. (b) Height.



(c) Angle. (d) Orientation.

Fig. 12: Experiment results of different relative positions.



(a) Different exterior interferences. (b) Results of different interferences.

Fig. 13: Verifying the impact of different exterior interferences.

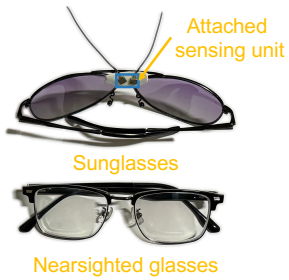


Fig. 14: Eyeglasses setup.

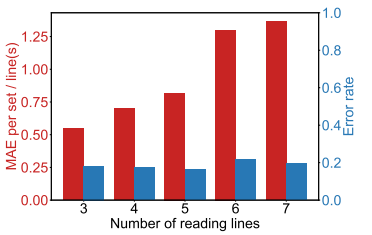


Fig. 15: Case study results.

the left of the participant; (ii) An interfering user walking 80 cm behind the participant; (iii) Background music increasing noise from 50 dB to 70 dB.

Fig. 13(b) shows that when the interfering user is active at at 80 cm, accuracy reaches 84% for saccades and 91% for blinks, indicating USEE effectively filters distant interference. However, at 40 cm, accuracy drops to 75% for saccades and 81% for blinks. Background noise has a negligible impact, demonstrating the system's ability to remove static components effectively.

6) Eyeglasses: We conduct experiments to explore the impact of wearing eyeglasses, including both nearsighted glasses and sunglasses, on the performance of USEE. The setup is shown in Fig. 14. The sensing unit is positioned directly in front of the participant's eyes at a distance of 10-15 cm, with each configuration repeated 10 times. When participants wear sunglasses, the average recognition accuracy is 82% for saccades and 84% for blinks. The relatively high accuracy is due to the sufficient transmission of ultrasonic signals through the gap between the lenses, demonstrating USEE's operational advantage over camera-based systems when users wear sunglasses. However, when participants wore nearsighted glasses, the accuracy decreased to 79% for saccades and 81% for blinks. The decline in accuracy is attributed to the attenuation of ultrasonic signal energy caused by the thicker lenses and the narrower gap between them, which limits signal propagation. To improve performance and evaluate the feasibility of integrating USEE into head-mounted devices, we affix the sensing unit to two types of eyeglasses, as shown at the top of Fig. 14. In this setup, the recognition accuracy significantly improved, reaching 88% for saccades and 90% for blinks. These results suggest the potential for USEE to be integrated as an intermediary layer in wearable devices.

E. Case Study

We further deployed USEE on the Raspberry Pi platform to demonstrate its potential as a real-life application in this context, *i.e.*, using USEE to estimate reading speed which is crucial in various human-computer interaction scenarios and also serves as a significant indicator of brain health [38]. We recruit two users to read the English text consisting of three to seven lines, with each line containing 14 to 18 words. Note that a saccade occurs when the user changes lines while reading, which serves as the basis for counting the number of reading lines. The Raspberry Pi platform is deployed above the display screen at a horizontal distance of 30 cm from the users' eyes and positioned vertically 5 cm above their eyes.

Two metrics are used to evaluate the experimental results: (1) The Mean Absolute Error (MAE) between the estimated and actual number of reading lines, which is used to evaluate the system's absolute accuracy; (2) The ratio of the MAE to the actual number of lines read by the user, employed to assess the system's stability. The results from Fig. 15 reveal that the MAE increases as the number of specified reading lines increases because the frequency of eye movements increases as the viewing content expands. When three lines of text are assigned, the MAE is 0.54 lines, and when it increases to seven lines, the MAE rises to 1.36 lines. Nevertheless, the average deviation rate of USEE remained relatively stable, staying around 18%, despite the increase in the actual number of lines. These observations validate the system's stability and demonstrate its potential for real-world applications.

V. Conclusion

We propose a method using ultrasound to detect extremely subtle eye movements with a single transceiver, including saccades and blinks, on both a research-purpose platform and a commodity Raspberry Pi. We model various interferences that occur during eye movement, and we design innovative signal processing methods to extract minute eye movement signals even in the presence of strong interference. Comprehensive experiments validate the effectiveness of USEE. We believe that USEE will introduce a new idea for instantaneous subtle

motion signals, offering new possibilities for detecting eye movements in various human-computer interaction scenarios.

ACKNOWLEDGMENT

We thank our anonymous reviewers for their insightful comments. This work is supported by the National Science Fund of China under grant No. 62272213.

REFERENCES

[1] P. Termsarasab, T. Thammongkolchai, J. C. Rucker, and S. J. Frucht, "The diagnostic value of saccades in movement disorder patients: a practical guide and review," *Journal of clinical movement disorders*, vol. 2, pp. 1–10, 2015.

[2] A. L. Boxer, S. Garbutt, W. W. Seeley, A. Jafari, H. W. Heuer, J. Mirsky, J. Hellmuth, J. Q. Trojanowski, E. Huang, S. DeArmond *et al.*, "Saccade abnormalities in autopsy-confirmed frontotemporal lobar degeneration and alzheimer disease," *Archives of neurology*, vol. 69, no. 4, pp. 509–517, 2012.

[3] N. Kosmyna, C. Morris, T. Nguyen, S. Zepf, J. Hernandez, and P. Maes, "Attentivu: Designing eeg and eog compatible glasses for physiological sensing and feedback in the car," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2019, pp. 355–368.

[4] L. Li, Y. Xie, J. Xiong, Z. Hou, Y. Zhang, Q. We, F. Wang, D. Fang, and X. Chen, "Smartlens: sensing eye activities using zero-power contact lens," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, 2022, pp. 473–486.

[5] J. Liu, D. Li, L. Wang, and J. Xiong, "Blinklistener: " listen" to your eye blink using your smartphone," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 2, pp. 1–27, 2021.

[6] H. Cheng, W. Lou, Y. Yang, Y.-p. Chen, and X. Zhang, "Twinkletwinkle: Interacting with your smart devices by eye blink," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 2, pp. 1–30, 2023.

[7] C. Quaia and L. M. Optican, "Three-dimensional rotations of the eye," *Adler's physiology of the eye: clinical application. New York: Mosby*, pp. 818–29, 2003.

[8] N. Pham, T. Dinh, Z. Raghebi, T. Kim, N. Bui, P. Nguyen, H. Truong, F. Banaei-Kashani, A. Halbower, T. Dinh *et al.*, "Wake: a behind-the-ear wearable system for microsleep detection," in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, 2020, pp. 404–418.

[9] S. B. Baray, M. U. Ahmed, M. E. Chowdhury, and K. Kise, "Eog-based reading detection in the wild using spectrograms and nested classification approach," *IEEE Access*, 2023.

[10] "Jins meme glasses," 2023. [Online]. Available: https://jinsmeme.com/en/

[11] E. Waisberg, J. Ong, M. Masalkhi, N. Zaman, P. Sarker, A. G. Lee, and A. Tavakkoli, "The future of ophthalmology and vision science with the apple vision pro," *Eye*, pp. 1–2, 2023.

[12] N. Kourkoumelis and M. Tzaphlidou, "Eye safety related to near infrared radiation exposure to biometric devices," *TheScientificWorldJOURNAL*, vol. 11, pp. 520–528, 2011.

[13] E. Allen, "Eye safety for proximity sensing using infrared light-emitting diodes photobiological effects of application note 1737 iec-62471: Photobiological," 2012.

[14] M. F. Harrison, J. P. Neary, W. J. Albert, U. Kuruganti, J. C. Croll, V. C. Chancey, and B. A. Bumgardner, "Measuring neuromuscular fatigue in cervical spinal musculature of military helicopter aircrew," *Military medicine*, vol. 174, no. 11, pp. 1183–1189, 2009.

[15] M. A. Hanson, *Health effects of exposure to ultrasound and infrasound: Report of the independent advisory group on non-ionising radiation*. Health Protection Agency, 2010.

[16] D. Aksu and M. A. Aydın, "Human computer interaction by eye blinking on real time," in *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE, 2017, pp. 135–138.

[17] X. Wang, J. Zhang, H. Zhang, S. Zhao, and H. Liu, "Vision-based gaze estimation: a review," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 316–332, 2021.

[18] J. H.-w. Hsiao, W. Liao, and R. V. Y. Tso, "Impact of mask use on face recognition: an eye-tracking study," *Cognitive Research: Principles and Implications*, vol. 7, no. 1, pp. 1–15, 2022.

[19] T. Akter, B. Dosono, T. Ahmed, A. Kapadia, and B. Semaan, "" i am uncomfortable sharing what i can't see": Privacy concerns of the visually impaired with camera based assistive applications," in *29th USENIX Security Symposium (USENIX Security 20)*, 2020, pp. 1929–1948.

[20] K. Li, R. Zhang, B. Chen, S. Chen, S. Yin, S. Mahmud, Q. Liang, F. Guimbretière, and C. Zhang, "Gazetrak: Exploring acoustic-based eye tracking on a glass frame," in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, 2024, pp. 497–512.

[21] S. Ramat, R. Leigh, D. Zee, A. Shaikh, and L. Optican, "Applying saccade models to account for oscillations," *Progress in brain research*, vol. 171, pp. 123–130, 2008.

[22] H. Wan, S. Shi, W. Cao, W. Wang, and G. Chen, "Resptracker: Multi-user room-scale respiration tracking with commercial acoustic devices," in *IEEE INFOCOm 2021-IEEE conference on computer communications*. IEEE, 2021, pp. 1–10.

[23] K. Sun, T. Zhao, W. Wang, and L. Xie, "Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proceedings of the 24th annual international conference on mobile computing and networking*, 2018, pp. 591–605.

[24] B. B. Basnet, P. K. Parajuli, R. K. Singh, P. Suwal, P. Shrestha, and D. Baral, "An anthropometric study to evaluate the correlation between the occlusal vertical dimension and length of the thumb," *Clinical, cosmetic and investigational dentistry*, pp. 33–39, 2015.

[25] G. W. McConkie, P. W. Kerr, M. D. Reddix, D. Zola, and A. M. Jacobs, "Eye movement control during reading: Ii. frequency of refixating a word," *Perception & Psychophysics*, vol. 46, no. 3, pp. 245–253, 1989.

[26] A. Mandel, S. Helokunnas, E. Pihko, and R. Hari, "Neuromagnetic brain responses to other person's eye blinks seen on video," *European Journal of Neuroscience*, vol. 40, no. 3, pp. 2576–2580, 2014.

[27] Y. Yao, M. M. H. Shandhi, J.-O. Hahn, O. T. Inan, R. Mukkamala, and L. Xu, "What filter passband should be applied to the ballistocardiogram?" *Biomedical Signal Processing and Control*, vol. 85, p. 104909, 2023.

[28] G. Ossberger, T. Buchegger, E. Schimback, A. Stelzer, and R. Weigel, "Non-invasive respiratory movement detection and monitoring of hidden humans using ultra wideband pulse radar," in *2004 International Workshop on Ultra Wideband Systems Joint with Conference on Ultra Wideband Systems and Technologies. Joint UWBST & IWUWBS 2004 (IEEE Cat. No. 04EX812)*. IEEE, 2004, pp. 395–399.

[29] T. Pozzo, A. Berthoz, and L. Lefort, "Head stabilization during various locomotor tasks in humans: I. normal subjects," *Experimental brain research*, vol. 82, no. 1, pp. 97–106, 1990.

[30] D. Da He, E. S. Winokur, and C. G. Sodini, "A continuous, wearable, and wireless heart monitor using head ballistocardiogram (bcg) and head electrocardiogram (ecg)," in *2011 Annual International Conference of the IEEE engineering in medicine and biology society*. IEEE, 2011, pp. 4729–4732.

[31] J. S. Stahl, "Amplitude of human head movements associated with horizontal saccades," *Experimental brain research*, vol. 126, pp. 41–54, 1999.

[32] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[33] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2011, pp. 4144–4147.

[34] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE transactions on signal processing*, vol. 62, no. 3, pp. 531–544, 2013.

[35] K. Ling, H. Dai, Y. Liu, A. X. Liu, W. Wang, and Q. Gu, "Ultragesture: Fine-grained gesture sensing and recognition," *IEEE Transactions on Mobile Computing*, vol. 21, no. 7, pp. 2620–2636, 2020.

[36] "Raspberry-pi-4-model-b," 2019. [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-4-model-b/

[37] S. Agarwal, D. Goel, and A. Sharma, "Evaluation of the factors which contribute to the ocular complaints in computer users," *Journal of clinical and diagnostic research: JCDR*, vol. 7, no. 2, p. 331, 2013.

[38] R. A. Armstrong, "Alzheimer's disease and the eye," *Journal of Optometry*, vol. 2, no. 3, pp. 103–111, 2009.