

論文 / 著書情報
Article / Book Information

題目(和文)	時間遅延がある視線連動型移動窓に対する操作主体感覚
Title(English)	Sense of Agency at a Temporally Delayed Gaze-Contingent Display
著者(和文)	Kim Junhui
Author(English)	Junhui Kim
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12903号, 授与年月日:2024年9月20日, 学位の種別:課程博士, 審査員:葭田 貴子,金子 寛彦,永井 岳大,小池 康晴,八木 透,吉村 奈津江, 横小路 泰義,赤間 啓之
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12903号, Conferred date:2024/9/20, Degree Type:Course doctor, Examiner:,,,,,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis



東京工業大学

令和 6 年度 博士論文

Sense of Agency at a Temporally Delayed Gaze-Contingent Display

指導教員

葭田 貴子 准教授

金子 寛彦 教授

提出者

機械系 ライフエンジニアリングコース

KIM JUNHUI

Abstract

The sense of agency (SoA), the subjective feeling of authoring one's actions and outcomes, is vital in human-computer interactions (HCI). While HCI involving hand and limb movements has been extensively researched, the domain of HCI with eye movements remains under-explored. This study demonstrates that temporal discrepancies between eye movements and corresponding interface updates disrupt SoA, subsequently affecting action planning. Moreover, the present study employs functional magnetic resonance imaging to reveal that the right angular gyrus (rAG) processes such disruptions. The research draws parallels between SoA's neural activation in hand or limb and eye movements and suggests that the eye, too, can be perceived as a causal initiator in HCI. Additionally, it highlights the unique impact of action-feedback mismatch in eye movements in HCI contexts, influencing attention deployment and action refinement.

Contents

Chapter 1	1
Sense of Agency.....	1
Internal Model.....	1
Remained Gaps	4
Aim of the Dissertation.....	7
Chapter 2.....	9
Method.....	13
Participants.....	13
Apparatus	14
Stimuli.....	14
Task and Procedure	18
Analyses.....	20
Results.....	24
Subjective Authorship	24
Categorical Response.....	24
Continuous Response.....	28
Behaviors	29
Response Time.....	29
Eye Movements	31
Contrast Analysis	33
Discussion	36
Sense of Agency.....	36
Behaviors	41
Critical Temporal Delay.....	46
Gaze-Contingent Display	48
Conclusion	48
Chapter 3.....	50
Materials and methods	51
Participants.....	53
Apparatus	54
Stimuli.....	54
Procedure	55

Analyses	56
Results	58
Subjective authorship	58
Behaviors	62
Discussion	66
Influences on the sense of agency	67
Influences on behaviors	68
Statistical attributes of jitter on the sense of agency	69
Gaze-contingent display	71
Conclusion	71
Chapter 4	73
Materials and Methods	75
Participants	76
Procedure and Task	76
Conditions	78
Stimuli	79
Apparatus and Magnetic Resonance Imaging Data Acquisition	80
Preprocessing of Functional Magnetic Resonance Imaging Data	81
Individual Analysis	82
Group Analysis	82
Delay-Modulated Analysis	83
Delay-Related Analysis	83
Incongruent vs. 0 ms Analysis	84
Behavioral Data Analysis	84
Results	86
Behavioral Results	86
Delay-Modulated Activation	88
Delay-Related Activation	90
Incongruent vs. 0 ms Contrast Activation	93
Discussion	95
Behaviors	95
Brain activations	96
Conclusion	99
Chapter 5	100

General Discussion	100
Limitations	101
Future Directions	101
Conclusion	104
Acknowledgements.....	106
References.....	108
Appendices.....	125
Appendix 1.1.....	125
Appendix 1.2.....	126
Appendix 1.3.....	128
Appendix 1.4.....	131
Appendix 2.1	132
Appendix 2.2.....	135
Appendix 2.3.....	138
Appendix 2.4.....	141
Achievements.....	143
Publication	143
International Conferences	143
Domestic Conference.....	144

Chapter 1

Sense of Agency

When engaging in everyday tasks, individuals naturally feel that they are performing the actions, often without giving this sensation much thought. For example, when entering a hotel, one might stand in front of an automatic door, expecting it to open. As the door opens, there is a perception that one's presence caused it to do so. This subjective experience, where one feels like the initiator of one's actions and outcomes, is known as the sense of agency [1]. This sense includes mechanisms that accurately differentiate the self from others when determining the authorship of an event [2,3]. A well-known phenomenon illustrating this concept is the inability to tickle oneself [4–7]. However, if there is a noticeable delay before the door opens, one might question whether they were responsible for triggering it or if someone at the front desk did it instead. Similarly, the self-tickling sensation is known to diminish when there is a temporal gap between the action and resulting sensation [8], indicating that the temporal discrepancy can influence the attribution of an event.

Internal Model

The “comparator model” has been proposed as an internal framework to explain the sense of agency [9–12] based on the efferent and afferent signals [9]. Based on this framework, when we intend to act, a signal is sent to the muscles, and a copy of this signal (efference copy) is stored to predict the expected outcome. The external event resulting from our action is then perceived by comparing this event (afferent signal) with the prediction made by the efference copy. Suppose the afferent signal matches the efferent copy. In that case, the model concludes that we are the event's initiators, leading to the cancellation of the sensation produced by our actions. This process of central cancellation has been used to explain various sensory attenuation phenomena, including the inability to tickle oneself [4–7], as well as findings in the visual domain, such as reduced sensitivity to self-generated

Gabor patches [10], diminished event-related potentials in response to self-initiated flashes [11], and the stability of visual perception despite rapid eye movements [12,13]. Conversely, any mismatch between the afferent and efferent signals indicates the possible involvement of another agent in causing the event. This comparison not only serves to cancel out sensations that are self-generated but also aids in identifying the trustworthy source of an event, with any discrepancies being used to adjust future motor commands for more accurate outcomes [14]. Deficiencies in this error correction process have been noted in individuals with schizophrenia [15]. Thus, the alignment between efferent and afferent signals is critical in shaping our perception of self-agency in external events.

The comparator model has gained support from empirical studies, including those investigating intentional binding [16] and the active rubber hand illusion [17], as well as experiments that systematically introduced a temporal delay between a participant's action and the resulting event [18–24]. In the active rubber hand illusion, participants saw a rubber hand placed on a box while their actual hand was hidden inside the box, covered up to the shoulder. A connecting rod attached the participant's index finger to the rubber hand's finger. Participants experienced a reduced sense of agency when the experimenter controlled the rod, causing both fingers to move, compared to when they controlled the movement themselves. The sense of agency further diminished when the rubber hand, detached and controlled by the experimenter, imitated the participants' finger movements with a temporal delay instead of when participants directly controlled the movements. Intentional binding, which refers to the perceived shortening of time between a voluntary action and its resulting event, was first observed by Haggard et al. [16] and has since been widely studied [25]. This phenomenon is explored through two paradigms: single-event time estimation and interval estimation. The method of the Libet clock, a variant of single-event time estimation, involves participants reporting the clock hand's position when they press a button or hear an auditory cue. In the

interval estimation method, participants report the perceived time interval between their action and the subsequent feedback [21,23]. Using the Libet clock, Haggard et al. [16] revealed that participants perceived the time interval as shorter when they voluntarily pressed the button, as opposed to when the action was involuntarily triggered by transcranial magnetic stimulation, underscoring the sense of agency. In studies that manipulated the temporal delay, participants' understanding of agency generally decreased when their actions and the visual feedback on a monitor were temporally misaligned [18–24].

The comparator model has faced criticism [26], being considered neither necessary nor sufficient to explain the sense of agency fully. Farrer et al. [19] observed that participants did not attribute actions to another agent even when there was a 1,000 ms delay between their actions and the visual feedback, suggesting that individuals might still perceive themselves as the cause of an event despite significant delays. This observation forms the basis of the first hypothesis: participants will continue attributing the cause of events to themselves, even when delays are present, especially when using eye movements to control stimuli. Another experiment that highlights the limitations of the comparator model is the “helping hand” experiment [27], where participants experienced a false sense of agency even when the action was actually performed by the experimenter. In response to these limitations, Synofzik et al. [26] proposed a two-component model for the sense of agency: one component is a non-conceptual feeling that identifies the initiator of an action, and the other is an explicit conceptual judgment that determines the authorship of an event based on contextual or conceptual factors like intentions and beliefs. In Farrer et al. [19], participants were informed that the experimenter might control the stimuli, leading them to believe they might not be in control. In the current study, participants were similarly instructed that the computer system might be responsible if they could not control the stimuli. By adapting the instructions and providing contextual cues, this approach leads to the hypothesis that the present study's

findings regarding the impact of delays on the sense of agency will mirror those of Farrer et al. [19]. The two-component model suggests that efferent-afferent congruency alone cannot fully explain the sense of agency. Based on this, Synofzik et al. [26] argued that explicit self-reports assess the judgment of agency, while Ebert and Wegner [21] proposed that intentional binding measures the implicit feelings of agency. Moreover, Wen et al. [23] demonstrated that temporal discrepancies between action and effect influence the explicitly reported degree of agency over an outcome but do not affect the binding effects. This study integrates these perspectives by employing rating questionnaires to gauge the intensity of participants' sense of authorship when using eye movements to control stimuli, hypothesizing that this sense will diminish in response to temporal delays. While recognizing the inherent limitations of questionnaires, they are valuable for assessing the impact of delay on the sense of agency.

Remained Gaps

Extensive research has been conducted on the sense of agency, yet studies explicitly addressing the role of eye movements in controlling human-computer interfaces (HCI) remain limited. In particular, the influence of action-outcome temporal discrepancies as an independent variable has not been thoroughly examined. While several studies have explored how individuals experience a sense of agency during social gaze interactions—such as perceiving that their gaze causes another person to follow it—and have analyzed the impact of delays in the other person's gaze response on this sense of agency [28–30], the application of temporal delays in feedback within a gaze-contingent paradigm for controlling computer interfaces is still largely unexplored. Research by Grynszpan et al. [31] and Gregori Grgič et al. [32] demonstrated that a sense of agency can indeed be elicited through eye movements using a gaze-contingent experimental setup. However, their exploration of temporal delays as an independent variable was limited. Gutzeit et al. [33] studied the reduction of intentional binding as temporal delays increased within a gaze-contingent paradigm, but their research focused on single saccades rather than continuous eye movements. This highlights a

significant research gap, indicating the need to investigate further how continuous eye movements—offering a higher degree of freedom—are influenced by temporal delays. Additionally, it is crucial to explore how users adapt their constant eye movements and employ behavioral strategies as their sense of authorship begins to diminish.

The limitations of relying solely on subjective reports to investigate the sense of agency underscore the need for additional methods to measure this phenomenon. As previously discussed, subjective reports have been criticized for capturing only a partial aspect of the sense of agency, specifically the conceptual judgment of agency [26]. The non-conceptual feeling of agency [26], also known as body agency [34], is proposed to be explained by the comparator model. Given these limitations, we used functional magnetic resonance imaging (fMRI) to capture the sense of agency more comprehensively. In particular, fMRI detects subtle changes in blood flow that correlate with brain activity, allowing researchers to identify which brain regions are responsible for specific functions. Although numerous studies have used fMRI to investigate the sense of agency [35–37] (for reviews, see [14,38]), these experiments have not included eye movements as an input method for human-computer interaction (HCI). In social gaze perception, research has examined brain activity related to how individuals perceive the direction of others' gaze [39,40]. However, the use of eye-gaze HCI in studies investigating the sense of agency has not yet been adopted. This gap in research highlights the need for further exploration into how eye movements as an HCI input can influence the sense of agency, particularly through neuroimaging techniques like fMRI.

The focus of this dissertation on the impact of temporal delays is significant not only from a theoretical perspective but also from a practical one. The sense of agency has been recognized as a crucial factor in evaluating user experience in human-computer interaction (HCI) [41], particularly concerning the user's feeling of control over the system. With the

emergence of eye-gaze as a new modality in HCI [42] and the increasing availability of accessible eye-tracking devices, such as Tobii [43] and the eye movement control features in Windows 10 and 11 [44], understanding how delays affect the sense of agency and user behaviors is becoming increasingly important. Despite this importance, current research has not adequately addressed practical applications, particularly concerning the range and variability of delays. First, gaze measurement and rendering delays can be substantial, reaching up to 500 ms in remote applications like remote surgery [45] and even several seconds in teleoperation over complex Internet connections [46,47]. Communication with orbiting space stations (e.g., the International Space Station) can face round-trip times of up to 1.2 seconds [48], and delays can extend up to 7 seconds when operating robotic arms on satellites [49]. Existing research on the sense of agency has primarily focused on delay ranges up to 1000-1100 ms [19,20,23], which is insufficient for addressing these practical scenarios. Second, teleoperated HCI often suffers inconsistent or “jittery” delays [50–56]. Teleoperation over a network involves round-trip time (RTT), the time it takes for a signal to travel to and from its destination [46]. RTT can vary due to several factors affecting packet delivery, including delays during transmission, propagation, queuing, and processing [47]. This variability is particularly relevant in teleoperation settings like remote surgery [45,57] and remote operation of satellites in low Earth orbit [58,59], where the HCI is prone to jittery delays. Despite its relevance, the inconsistency in delays has not been considered in studies on the sense of agency. This gap highlights the need for further research that accounts for the range and variability of delays in practical HCI applications.

In eye-gaze HCI, specifying the type of agency referenced, especially when highlighting its significance, is essential. The agency associated with the classical comparator model in visual stability research should be differentiated from the type of agency encountered in scenarios such as waiting for a coffee machine, gaze typing, or button

pressing in eye-gaze systems initially designed for patients with (Amyotrophic Lateral Sclerosis) ALS, where longer delays may be tolerated. The core of agency in gaze-contingent systems centers on immediate, continuous interaction; any substantial delay disrupts the user's experience. Even a brief delay in these systems can completely shatter the immersive, first-person experience crucial to effective eye-gaze control. While some might argue that a belief-based sense of agency suffices to evaluate HCI quality, extended delays alter the interaction model from real-time, natural control to one requiring more deliberate, anticipatory actions. This shift undermines the intuitive, direct control that users expect. As seen in first-person shooter (FPS) video games, minimal delay is critical to maintaining a fluid and immersive experience [60]. Similarly, in gaze-based interfaces, even minor temporal disruptions can break the natural flow of interaction, causing the system to feel less responsive and significantly reducing user satisfaction.

Aim of the Dissertation

This research addresses the identified gap by investigating how delayed feedback in eye-gaze HCI influences self-reported subjective sense of agency, eye movements, and brain activity measured by fMRI. With the fMRI, the third experiment aimed to capture the brain activity directly related to the reported sense of agency while also investigating the activity associated with the amount of the action-effect delay. This examined whether the activations could be related to the comparator model.

This dissertation is particularly interested in whether different modalities, such as eye movements versus limb movements, lead to distinct dynamics on the impact of the delay. Given that the eyes were not initially intended for direct manipulation of computer interfaces, this unique characteristic of the eye-gaze modality might yield different responses to temporal delays. Suppose these findings are consistent with previous studies. In that case, they will help expand the current models of the sense of agency. Furthermore, the current

investigation of the impact on the participant's behavior allows detailed insights related to the comparator model. How perceived temporal mismatch (i.e., delay) refines action to achieve the goal. Deviating eye movements were assumed to be derived by strategies adjustment after a mismatch perception.

Extended temporal delays up to 4 s were adopted for the practical gap in the first experiment. In the second experiment, the impact of the jitter on the sense of agency was investigated by comparing inconsistent to consistent delays. Additionally, I sought to suggest guidelines in the delay ranges where the sense of agency is maintained. The current study aimed to identify a critical temporal delay—a threshold at which a user's experience becomes notably disrupted in eye-gaze HCI. Furthermore, this dissertation aimed to suggest deviating patterns of eye movements as a measurable proxy of the user's hindered experience. Further, investigating the underlying brain activations (third experiment) may shed light on future practical research about where the eye-gaze sense of agency would be measured and how the activation varies concerning the sense of agency.

I developed a novel, simple visual task with a gaze-contingent display that allows us to serve the present study's aim. I first investigated the effect on subjective reports and eye movements (Experiment 1). Then, the investigation was further extended, with inconsistent delays (Experiment 2). Finally, I examined the delay-modulated brain activation with an fMRI (Experiment 3).

Chapter 2

Temporal delays ranging from 0 to 4000 ms were introduced to explore their effects in teleoperation contexts, as such extensive delays have not been previously examined in these scenarios. The hypothesis posits that the impact of action-feedback temporal delay on the sense of agency in eye-gaze human-computer interaction (HCI) will resemble the patterns observed in limb-based modalities. Specifically, it is anticipated that participants will attribute events to themselves, rather than to the other agent (i.e., the computer), when temporal delays are present in eye-gaze HCI, aligning with prior research [18]. Furthermore, the study aims to evaluate the strength of the participants' sense of authorship, hypothesizing that it will weaken as temporal delays increase, consistent with findings from studies on limb modalities [17,20–23]. This hypothesis is grounded in the understanding that the predictive mechanism involving the efference copy is relevant not only to limb and foot movements [6,7,52–54] but also to eye movements (i.e., corollary discharge), which is crucial for maintaining visual stability during rapid gaze shifts [5,12,13,55]. Evidence for this was provided in a study where disrupting the corollary discharge circuit in monkeys impaired visual stability [56]. Similarly, Lindner et al. [57] observed that schizophrenia patients exhibited a diminished ability to distinguish their own smooth eye pursuits from perceived motion.

Temporal delays ranging from 0 to 4000 ms were introduced, as such extended delays have not been previously explored in teleoperation scenarios. The hypothesis is that the effect of action-feedback temporal delay on the sense of agency in eye-gaze HCI will mirror the patterns observed in limb modalities. First, it is expected that the participants identify themselves, rather than the other agent (i.e., the computer), as the origin of events when temporal delays occur in eye-gaze HCI, consistent with earlier findings [19]. Second, I aim to assess the strength of their sense of authorship, hypothesizing that it will diminish with

increasing temporal delays, similar to what has been observed in studies involving limb modalities [18,21–24]. This hypothesis is based on the understanding that the predictive mechanism involving the efference copy applies not only to foot and limb movements [4,5,8,61,62] but also to eye movements (i.e., corollary discharge) in maintaining visual stability despite rapid gaze shifts [9,12,13,63]. This was demonstrated in a study where inactivating the corollary discharge circuit in monkeys disrupted visual stability [64]. Similarly, Lindner et al. [65] found that schizophrenia patients had a reduced ability to distinguish their own smooth eye pursuit from motion perception.

In designing the current experiment to assess the sense of agency, I did not use the intentional binding paradigm. Since one of the objectives was to extend the findings to teleoperating HCI settings, a continuous task involving eye movements was implemented, with participants reporting their sense of agency. It was also particularly cautious of the complexities brought about by the chronostasis effect, which can complicate experimental design. As described by Yarrow et al. [66], chronostasis refers to the perceived time expansion immediately after a saccade. This effect poses challenges in a task simplified to a gaze-triggered button click, which is a common setup in intentional binding studies [25]. Instead, I opted for a questionnaire-based approach. Although questionnaires are criticized for capturing only the judgment of agency rather than the full experience, they still provide enough information to evaluate the current hypothesis. This approach also serves as a practical starting point for investigating this previously unexplored area, enabling us to compare these findings with those of prior studies.

For the first hypothesis, I adapted Farrer et al.'s [19,20,67] method of assessing agency by providing participants with three response options: 'the eye-gaze HCI's updates aligned my eye movements' (self), 'aligned my eye movements but with a delay' (delay), and 'the computer system controlled the eye-gaze HCI' (other). In this study, the original 'partial

control' option from Farrer et al. [19,20,67] was modified to 'delay' to better suit the controllable aspect of the current HCI task. It has been suggested that the binary choices of “self” and “other” do not fully encompass situations where participants feel their actions contributed to the outcome, albeit not completely or directly (e.g., “I felt that I influenced the outcome, but the causal link between my actions and the result was temporally distorted”). Therefore, I introduced the “delay” option to account for cases where participants might still experience sense of authorship despite temporal delays affecting the interface’s responsiveness. In addition, as suggested by Farrer et al. [67], an uncontrollable condition was implemented to determine if this questionnaire could detect participants attributing control to another agent. By adapting Farrer et al.’s [19,20,67] agency measurement method, I aimed to replicate their established approach and anticipated that the questionnaire in this study would yield comparable results within the context of eye-gaze HCI.

For the second hypothesis, a rating questionnaire was implemented to specifically assess the altered sense of agency within the range of identifying ‘I’ as the event author, excluding the “other” option. The question utilized a six-point scale, ranging from 6 (full control: ‘I fully manipulated the window’) to 1 (no control: ‘I couldn’t possibly manipulate the window’). While the questionnaire is similar to those used in previous studies [18,21–23], it offers greater clarity by avoiding ‘disagree’ responses, which could be misinterpreted as indicating control by another agent. Instead, this rating scale guides participants to exclude the “other” consideration when rating their level of authorship. By omitting the “other” category, I aimed to more reliably capture the nuanced variations in participants’ perception of themselves as the author. Although the “delay” response in categorical questionnaires represents altered self-authorship, there is a high likelihood that “delay” response’s proportion will converge at lower temporal delays (around 300 ms as suggested by prior research [19]) and remain consistent across longer delays. This can lead to an

oversimplification of participants' perceptions in extended delay scenarios. Continuous scales offer finer detail by measuring how the perception of 'I' as the author diminishes across varying delay conditions beyond 300 ms. By utilizing these two questionnaires, current study propose that both distinct changes in altered 'I' authorship and subtle shifts in attenuated 'I' authorship within the context of eye-gaze HCI, can be capture.

It is also hypothesized that participants will adjust their strategies in response to these delays. This hypothesis aligns with the comparator model, which suggests that differences between expected and actual outcomes are used to refine motor commands to complete tasks [14]. In this study, participants performed a visual search using the current eye-gaze HCI, which required them to search in a sequential manner. The current HCI aimed to disrupt their behavior through delays. In this HCI, characters are readable only within a controllable window that tracks the user's eye movements, while the surrounding area is blurred. The disruption in visual search with the HCI is a crucial factor to consider in practical applications of this HCI. Through the current measurements using two questionnaires, I aimed to determine a critical temporal delay—a threshold where the user's experience becomes significantly disrupted in eye-gaze HCI. Previous research has shown that participants begin to detect visual artifacts in eye-gaze HCI at delays of 50–70 ms [68,69]. However, the threshold at which users' sense of authorship becomes disrupted may differ.

In eye-gaze human-computer interaction (HCI) research, gaze-contingent multi-resolution rendering [70], also known as "foveated imaging" or "foveated rendering" [68,71–74], is utilized. This technique generates frames with multiple-resolution regions and smoothly blended edges, taking into account the human eye's visual acuity gradient from the fovea to the peripheral retina [75]. Foveated rendering optimizes data processing efficiency while maintaining the user's visual experience [76], and it is commonly applied in virtual reality headsets, which require demanding rendering cost with high display resolutions [72–

74,77]. In this experiment, a simple bi-resolution gaze-contingent display with high-resolution and blurred regions was used, featuring an unblended border in contrast to other studies [76,78–80]. This design was intended to highlight differences in saliency with the current HCI, offering valuable insights for optimizing interface design and improving user experience across diverse HCI settings. Beyond examining the effects of temporal delays, variations in window saliency (i.e., visibility) were introduced. Initially, a high contrast was created between the window and its surrounding area. Later, the design was adjusted to more closely align the visibility of these elements. This methodological choice, incorporating two distinct visibility levels, was intended to support the broader applicability of the findings.

Method

Participants

Due to the presence of Chinese characters in the stimuli, only individuals capable of reading Chinese characters were recruited. In total, 11 participants were initially recruited; however, three were excluded during the experiment as their calibration accuracy exceeded 1°. Data from eight subjects (average age: 24.4 years; standard deviation [SD]: 1.5; range: 22–27 years; one woman and seven men: two Koreans, five Japanese, and one Chinese) were examined. All participants were students (two undergraduates and six master's students) from the Tokyo Institute of Technology. Undergraduate students were informed of the study in their classes, while master's students were notified via their advisors through emails sent by the experimenters. Each participant received a 1,000-yen book voucher card for each hour of participation. All participants had normal or corrected-to-normal vision. Written informed consent was obtained from all participants prior to the study, and the research was approved by the Tokyo Institute of Technology Human Subjects Research Ethics Review Committee. Data collection occurred from October 6, 2020, to January 26, 2021, followed by statistical analyses and evaluations aimed at addressing the research questions. It is acknowledged,

following Lakens [81], that no formal sample size justification was provided. The sample size was constrained by available resources, as each participant needed more than two months to complete 60 sessions, making it difficult to recruit additional participants.

Apparatus

The experiments were conducted in a dimly lit room to help participants reposition their hands during responses. Participants' heads were stabilized using a chin rest throughout the experiment. Visual stimuli were presented on a 24-inch LCD monitor (BenQ Corp., XL2411t) with a resolution of 1920×1080 , a frame rate of 144 Hz, and a 60 cm viewing distance. The monitor's input lag, given the 144 Hz refresh rate, is theoretically 7 ms. Monocular eye movements (left eye) were tracked using an EyeLink 1000 Plus eye tracker (SR Research Ltd.), which features a spatial resolution of less than 0.02° and a sampling rate of 2,000 Hz. The average sample delay of the eye tracker was less than 1.4 ms, with a standard deviation of under 0.4 ms. The minimum temporal delay from an eye movement to the corresponding feedback on the monitor includes the eye tracker's sample delay, the experiment software's time duration of processing (which updates every 1 ms), and the monitor's input lag. Participants provided their responses using a numeric keypad on a Lenovo SK-8825 keyboard. The software for the current experiment, a modified version of the GCWINDOW template from SR Research Ltd., was specially designed to collect eye movement data, buffer it, and update the stimuli with a temporal delay. During each session, the software stored eye movement data from one particular trial for later replay. Frames were rendered every 7 ms, while data acquisition and internal computations were performed every 1 ms. Calibration and validation were conducted using a nine-point method, with a validated accuracy of less than 1° .

Stimuli

In the present eye-gaze HCI setup, peripheral blurring was applied during participants'

visual search, allowing Chinese characters to be readable only within a configurable window. It was aimed to require participants to search sequentially. This design was intended to effectively disrupt their behavior with delays. However, an inherent concern in the current HCI is that saccade amplitudes have been shown to decrease with peripheral blurring [76,78–80,82]. For example, Reingold and Loschky [76] found that participants exhibited delayed first saccades and reduced saccade amplitudes when using a peripheral blur window while tracking a moving object in a video. It has been suggested that these effects on eye movements are due to the reduced saliency of peripheral targets [83]. These observations are relevant to the current study for two reasons: first, many dominant eye-gaze HCI systems utilize peripheral blurring [68,71–74], and second, the effect of the window was considered to be outside the scope of the present study.

To counter the decreasing saccade amplitudes, which is the unintended effect, a distinct stimulus was created. Despite the expected decrease in visibility within the blurred periphery of the current stimuli, a lesser degree of reduction compared to natural stimuli was anticipated, thereby reducing the impact of peripheral blurring. A visual search array (Figure 1.1A) containing nine black Chinese characters (4.71 ± 1.32 cd/m² of luminance) was presented in a 3×3 virtual matrix on a gray background (30.41 ± 3.15 cd/m² of luminance). Each character measured 3° in size, and the horizontal and vertical spacing between adjacent characters was 11° .

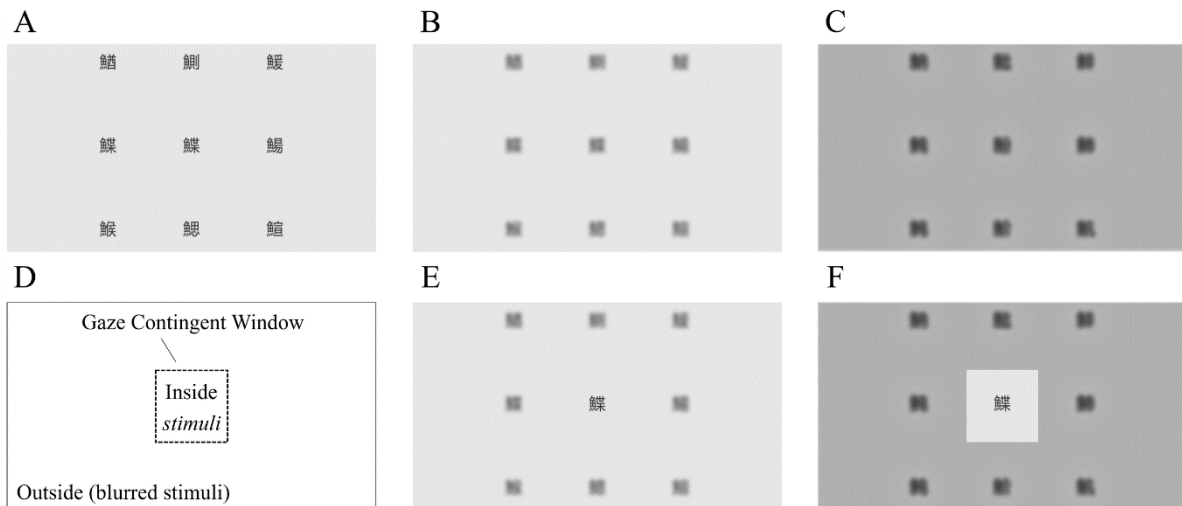


Figure 1.1. A) An example of an image generated through random image generation; a total of 540 images were produced. B) One image was randomly selected and subsequently blurred. C) In addition to applying blurring, the image was also darkened. D) A virtual square window was positioned at the participant's fixation point, displaying a clear image within the inner region of the window, while the outer region showed a blurred image. E) An example of the stimulus with the blurred image (Figure 1.1B) is shown when the participant fixated on the central Chinese character (low visibility condition). F) An example of the stimulus with the blurred and darkened image (Figure 1.1C) is presented when the participant fixated on the central Chinese character (high visibility condition). Reproduced from Kim et al. [84].

The condition with low visibility was created by applying a blur to the periphery, while the condition with high visibility involved both darkening and blurring the periphery. Two types of blurred images were utilized: one where only blurring was implemented (Figure 1.1B) and another where both darkening and blurring were applied (Figure 1.1C). To blur the images, a Gaussian filter was applied to one randomly chosen image from the set of 540 (Figures 1.1B and 1.1C) with Adobe Photoshop 2020. The low-pass filter had a standard deviation of 10 pixels ($.2^\circ$), attenuating the signal by 3 dB at a spatial frequency of 0.57 cycles/ $^\circ$, with further attenuation at higher frequencies. The blurred Chinese characters showed increased luminance (15.27 ± 3.76 cd/m ^{2}), while the gray background luminance

remained constant. The darkening was achieved by reducing the red-green-blue (RGB) brightness values (0–255 for each) of the blurred image by 54 units per channel. As a result, the RGB brightness of the blurred image decreased from 231 to 177, resulting in a background luminance of 20.28 ± 4.29 cd/m² and a character luminance of 5.21 ± 2.04 cd/m².

A 10° square-shaped window was dynamically aligned with participants' fixation points, tracking their eye movements in real-time (Figure 1.1D). The window's edges extended 5° horizontally or vertically from the fixation position. The array of Chinese character (Figure 1.1A) appeared within the inner region of the gaze-contingent window, while one of the blurred images (Figures 1.1B and 1.1C) was displayed in the surrounding outer area (Figure 1.1D), creating a clear boundary between the two zones with a step transition. An instance of a displayed image featuring a blurred surrounding area is depicted in Figure 1.1E (low visibility), and another example with both a blurred and darkened surrounding area is shown in Figure 1.1F (high visibility). Although the display was designed so that only one character could be clearly inspected at a time when fixated upon, participants had the potential to position their gaze between two characters to view both simultaneously. However, the current analysis indicated that participants did not adopt this strategy (see Appendix 1.4).

Chinese characters were selected based on two specific criteria (rather than using the English alphabet; see Appendix 1.3 for a pilot study comparing English and Chinese characters). First, each character had to be a radical related to a fish name, ensuring that all characters belonged to the same semantic category, thereby minimizing meaning- or category-based attentional biases. Second, the nine characters chosen were required to have identical stroke counts, ensuring they shared similar spatial frequency, contrast, complexity, and visibility, particularly in peripheral vision. 239 Chinese characters (listed in Appendix 1.1) were selected for this study.

For the visual search array (Figure 1.1A), a 50% probability was set for the presence of a target character among the eight surrounding locations, resulting in two types of arrays: target and non-target. This uncertainty was introduced to prevent participants from predicting the target's location prior to completing the visual search, which could potentially influence their behavior. The lowest correct response rate for individual tasks was 96%, suggesting that participants engaged with each trial without relying on prediction (see Appendix 1.2 for response rates by participant and condition).

Using preprogrammed random image-generating software in Unity, 540 image files were created. Nonetheless, because the images were not sampled evenly, the probability of the target's presence and its location varied. Among the 540 images, 272 were arrays without targets. Among the 268 target arrays, the distribution of target appearances across the eight possible locations (starting from the middle top and moving clockwise) was as follows: 34, 29, 32, 32, 35, 40, 37, and 29.

Task and Procedure

Participants were instructed to locate a target character that matched the one displayed at the center of the screen within the eight peripheral locations. When the target character was present, they were to press the corresponding button on the numeric keypad; if the target was absent, they were to press '0'. Participants were informed that the gaze-contingent window would occasionally have a delay following their eye movements and might sometimes be controlled by the computer system. They were instructed to perform the visual search even when they could not control the window and to press '0' if they could not find the target during these uncontrollable trials. The task concluded when participants confirmed their response by pressing the Enter key.

Two questionnaires were asked to complete after each task. First, they answered a categorical authorship questionnaire, indicating whether they perceived that they controlled

the movements of the window with their gaze (self), noticed a temporal distortion or delay in the causal relationship (delay), or perceived that the movements were controlled by the computer rather than their gaze (other). Additionally, they completed an authorship rating questionnaire, where they rated the degree to which they felt their eye movements caused the window to move. This was done on a six-point scale, with six indicating "I fully manipulated the window" and one indicating "I could not possibly manipulate the window."

As illustrated in Figure 1.2, each trial began with a drift check. The visual search commenced after a blurred image was displayed for 4,680 ms. This duration was chosen to ensure that the experimental software had enough time to buffer a participant's gaze behavior during delayed trials. The 4,680 ms duration was consistent across all conditions, and participants were instructed to fixate on the center character during this period. After completing the visual search, participants then answered the two authorship questionnaires.

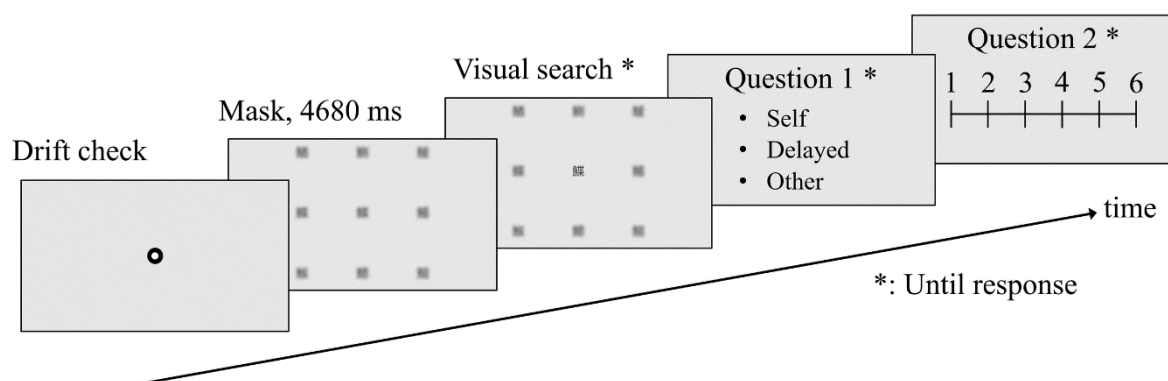


Figure 1.2. The trial sequence: following the drift check, participants waited until the gaze-contingent window become visible. After providing their response, they were required to complete two questionnaires. The drift checks and questionnaires are depicted with exaggerated sizes for explanatory purposes. Reproduced from Kim et al. [84].

Each participant completed 60 sessions (1,080 trials), containing two visibility conditions, with 30 sessions (540 trials) for each condition. Participants first completed all 30 sessions under one visibility condition before switching to the other, with the order of the

conditions counterbalanced across participants. Each session consisted of 18 trials: two practice trials, 15 delayed trials, and one playback trial. The 15 delayed trials and the playback trial were randomly sequenced following the two practice trials. The practice trials were identical to the 0 ms delay trials, meaning they introduced no intentional temporal discrepancy.

In the delayed trials, the gaze-contingent window was subject to one of 15 relative temporal delays: 0, 30, 60, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, or 4,000 ms. The playback trial represented an uncontrollable condition in which the window replayed the participant's recorded gaze behavior from one of the two practice trials, not corresponding to their real-time gaze. Between the two visibility conditions, participants completed two sessions (36 trials each) without the gaze-contingent display, referred to as the "no window" condition, to assess the influence of the window on participants' behaviors. With this condition, participants were not required to answer any subjective questionnaires.

During each "no window" session, participants were presented with 36 randomly selected images from a set of 540 images. The same set of 540 images was presented twice within each visibility session, in the same sequence, to aid in data preprocessing. Participants took breaks, recalibrated, and underwent revalidation after each session. Throughout the experiment, I periodically asked participants if they remembered any of the stimuli. As none of the participants reported any recollection, and considering the counterbalanced sequence between participants, it is concluded that reusing the stimuli in the same sequence across the visibility conditions did not affect the results.

Analyses

A significance threshold of 0.05 was applied to all analyses to determine statistical relevance. To test the first hypothesis, which examines whether participants correctly identified the action agent, categorical authorship responses were analyzed. Two generalized

linear mixed models (GLMM; SPSS 24.0) were employed, excluding playback trials, to compare "self" versus "delay" responses and "other" versus "delay" responses, by implementing a generalized logit model with a multinomial probability distribution. For the random effects, participants were modeled with variance components as the covariance structure. Window visibility and temporal delay were treated as fixed effects. Due to data loss in the subjective reports, 42 responses were excluded, resulting in 7,158 responses being analyzed. The Satterthwaite Formula was used to approximate the degrees of freedom, given the relatively small sample size and the unbalanced data resulting from partial data loss.

To explore the second hypothesis, which posits a decreasing intensity of the sense of authoring with increasing delay, authorship ratings were analyzed. A repeated-measures ANOVA with a visibility $\times$ delay design was conducted on the participants' mean authorship ratings for the delayed window condition, applying the Greenhouse-Geisser correction. To assess the reduced ratings in the playback condition, the two visibility data groups were combined, as neither visibility nor its interaction showed a significant main effect (as detailed in the Results section). A paired t-test was then performed to compare participants' mean rating scores between the playback and the 4,000 ms delay conditions, with each group containing 16 data samples.

For the third objective, which focused on the impact of delays on participants' behavior, response times were analyzed to determine whether the current eye-gaze HCI compelled participants to perform a visual search in a serial manner, thereby disrupting their behavior due to delays. In this study, response times were expected to be shorter when target stimuli were present compared to no-target stimuli. The time from stimulus onset to the registration of the participant's final input was measured, with all other inputs excluded, as responses were often re-registered when participants needed to reposition their hands on the keypad. A repeated-measures ANOVA with a visibility $\times$ target $\times$ delay design was applied

to the participants' mean response times under the delayed window condition, using the Greenhouse-Geisser correction.

In addition, the impact of the gaze-contingent window on participants' behavior was investigated. Given that visibility did not show a main effect (as detailed in the Results section), the high- and low-visibility groups were combined into one and a repeated-measures ANOVA with a target $\times$ window presence design, applying the Greenhouse-Geisser correction, was conducted. The four data groups were included in this analysis: 1) no window with no target stimuli, 2) no window with target stimuli, 3) combined data from high/low-visibility windows with 0 ms delay with no target stimuli, and 4) combined data from high/low-visibility windows with 0 ms delay with target stimuli. Each window-presence group comprised 16 data samples, while each window-absence group contained eight. In this ANOVA, since the effect of the target has already been discussed, only the effect of window presence was reported.

The effect of delay on eye movements were investigated, specifically focusing on the relative distribution of fixation duration and saccade amplitude. Saccades and fixations were identified from raw gaze data using SR Research Data Viewer software. The predefined minimum thresholds for detecting saccades were set at 0.1° , $30.0^\circ/\text{s}$, and $8000.0^\circ/\text{s}^2$. Fixations that occurred outside the stimuli, those starting immediately before or after a blink, and fixations shorter than 120 ms were omitted from the analysis. The mean relative frequencies of fixation durations were calculated based on the representative frequencies of the eight participants. For every participant and across all conditions, the frequency of fixation durations within each 25 ms bin in the 0–800 ms range was normalized to a relative frequency. This normalization was necessary because the number of eye movements varied significantly among participants due to individual differences in response times. Similarly, the mean relative frequencies of saccade amplitudes were calculated using the representative

frequencies of the eight participants. For each participant, the frequency of saccade amplitudes within each 0.5° bin in the $0\text{--}15^\circ$ range was normalized to a relative frequency for each condition.

Additional analyses were conducted to investigate changes in the proportion of "self" responses in the categorical questionnaire and behavioral changes in eye movements within the $0\text{--}500$ ms range (referred to as contrast analysis), prompted by observed variations in this range (see the Results section). The $200\text{--}300$ ms range was specifically chosen because the proportion of "self" responses dropped below 25% and then stabilized. The rating scores were not included in the contrast analysis, as the focus was on understanding the significant shifts in participants' experiences within this critical $200\text{--}300$ ms delay. The reported mean rating scale gradually decreased across all delay conditions, with greater emphasis on exploring the overall variation throughout the delays rather than on the changes within the $200\text{--}300$ ms range.

With the observation of a shifting mode in the fixation duration distribution as delay increased, the x-axis location of this mode was examined. For each participant and condition, fixation duration distribution was calculated using 50 ms bins within the $0\text{--}800$ ms range. Then the bin where the mode peaked was identified and was refined its position in 1 ms increments, locating the maximum count within a -50 to 50 ms range from the original position. Additionally, increase in saccade amplitude under 1.5 degrees with the delay was investigated. For each participant and condition, the cumulative distribution for saccades at 1.5 degrees was calculated based on the previously determined distribution using 0.5° bins.

Next, the changes in each value—proportion of "self" responses, the location of the shifting mode in fixation duration, and the cumulative distribution for saccades at 1.5 degrees—were calculated for each 100 ms increase in delay (i.e., $0\text{--}100$ ms, $100\text{--}200$ ms, $200\text{--}300$ ms, $300\text{--}400$ ms, and $400\text{--}500$ ms). Two multivariate ANOVAs were conducted:

one analyzing delay $\times$ visibility and the other examining delay increase $\times$ visibility. Post-hoc pairwise comparisons were performed using the Bonferroni correction, and significant differences found in the post-hoc analysis were plotted using Compact Letter Display (CLD) graphs.

To investigate the critical temporal delay based on the categorical authorship questionnaire, the proportion of "self" responses across each delay and visibility condition was analyzed. A delay $\times$ visibility repeated-measures ANOVA with Greenhouse-Geisser correction was performed. Since the effects of delay and visibility have already been discussed, only the results of the post-hoc pairwise comparisons will be reported, which were conducted using the Bonferroni correction. For the authorship ratings, post-hoc pairwise comparisons with Bonferroni correction were also conducted after the visibility $\times$ delay repeated-measures ANOVA described earlier. Similarly, for response time, post-hoc pairwise comparisons with Bonferroni correction were performed after the visibility $\times$ target $\times$ delay repeated-measures ANOVA, as previously detailed.

Results

Subjective Authorship

Categorical Response

Categorical authorship variable across three levels were assessed: self, delay, and other. The results are illustrated in Figure 1.3.

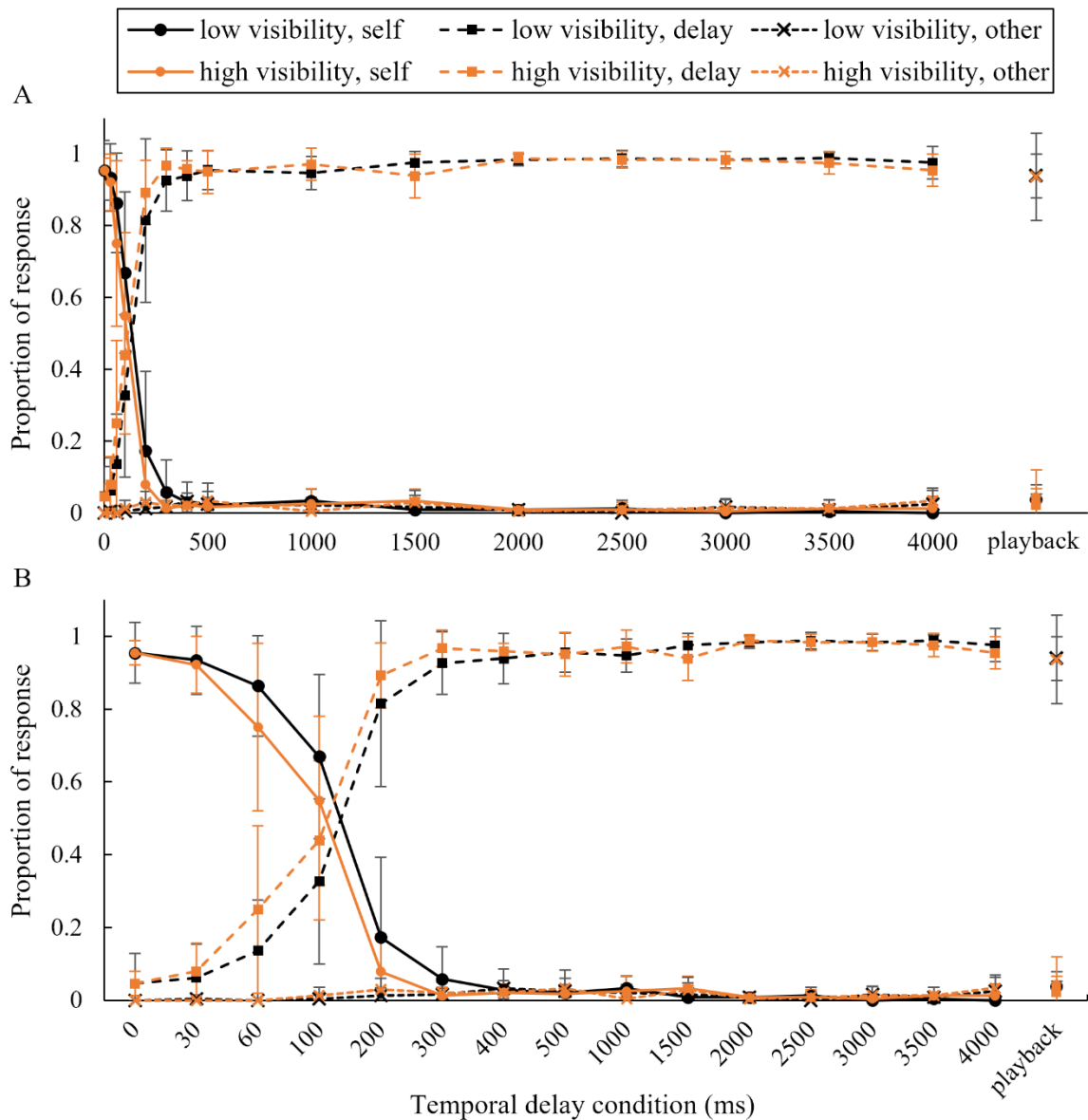


Figure 1.3. The mean proportions of reported authorship as self, delay, or other. Means and standard deviations were calculated across participants for each condition and window visibility. The results are displayed using both a linear x-axis (A) and a nonlinear x-axis (B), the latter chosen for comparison with a previous study where a similar categorical response was reported on a nonlinear x-axis [19]. Reproduced from Kim et al. [84].

With the GLMM comparison between the “self” and “delay” response, a significant main effect of delay was observed ($b = -5.770$, $SE = 1.846$, $t = -3.126$, $p = .002$, Odds Ratio

[OR] = .003), indicating that participants were less likely to choose the “self” response as the delay increased. When comparing the low visibility to the high visibility window condition (LV – HV), window visibility also showed a significant main effect ($b = 1.175$, $SE = 0.378$, $t = 3.111$, $p = .002$, $OR = 3.238$), suggesting that the “self” response was more frequently chosen with a low visibility window. As illustrated in Figure 1.3, the proportion of “self” responses decreased with longer delays in the low visibility window than in the high visibility window, indicating that participants detected delays at shorter intervals in the high visibility window. This pattern is further supported by a significant interaction between Delay and LV – HV ($b = -6.223$, $SE = 1.965$, $t = -3.168$, $p = .002$, $OR = 0.002$). Table 1.1 detailed all covariate parameters and fixed coefficients from the GLMM analysis.

Table 1.1. Fixed coefficients and estimated covariance parameter from the GLMM comparing the “self” to the “delay” response.

Covariance parameter	Estimate	Standard error	Z score	P value
Participants	0.181	0.104	1.744	.081
Fixed effects	Coefficient	Standard Error	T score	P value
Intercept	0.758	0.264	2.870	.006
Delay	-5.770	1.846	-3.126	.002
Low visibility vs. high visibility (LV – HV)	1.175	0.378	3.111	.002
Delay $\times$ LV – HV	-6.223	1.965	-3.168	.002

With the GLMM comparison between the “other” and “delay” response, the main effect of delay was not significant ($b = -0.068$, $SE = 0.213$, $t = -0.318$, $p = .750$, $OR = 0.934$), indicating that the “other” response was rarely selected in the delayed condition. Similarly, the comparison between the low visibility and high visibility window (LV – HV) showed that window visibility was not a significant main effect ($b = -0.088$, $SE = 0.296$, $t = -0.296$, $p = .767$, $OR = 0.916$), suggesting that the “other” response was infrequently chosen across both window conditions. Additionally, the interaction between Delay and LV – HV was not significant ($b = -0.080$, $SE = 0.248$, $t = -0.322$, $p = .747$, $OR = 0.923$). Table 1.2 detailed all covariate parameters and fixed coefficients from the GLMM analysis.

Table 1.2. Fixed coefficients and estimated covariance parameter from the GLMM comparing the “other” to the “delay” response..

Covariance parameter	Estimate	Standard error	Z score	P value
Participants	0.254	0.194	1.312	.190
Fixed effects	Coefficient	Standard Error	T score	P value
Intercept	-3.903	0.325	-12.021	.000
Delay	-0.068	0.213	-0.318	.750
Low visibility vs. high visibility (LV – HV)	-0.088	0.296	-0.296	.767
Delay $\times$ LV – HV	-0.080	0.248	-0.322	.747

A delay $\times$ visibility repeated-measures ANOVA focusing solely on the “self” response, was conducted to investigate the critical temporal delay based on the authorship

questionnaire. Since the effects of delay and visibility have already been discussed, only the post-hoc pairwise comparisons with Bonferroni correction was reported. The proportion of “self” responses was significantly different compared to the 0 ms delay ($M = .954$) when the delay reached 200 ms ($M = .126$, $p < 0.001$) and higher ($p < 0.001$).

Continuous Response

The authorship rating was assessed using a 6-point scale to measure the degree of authorship. The results are displayed in Figure 1.4.

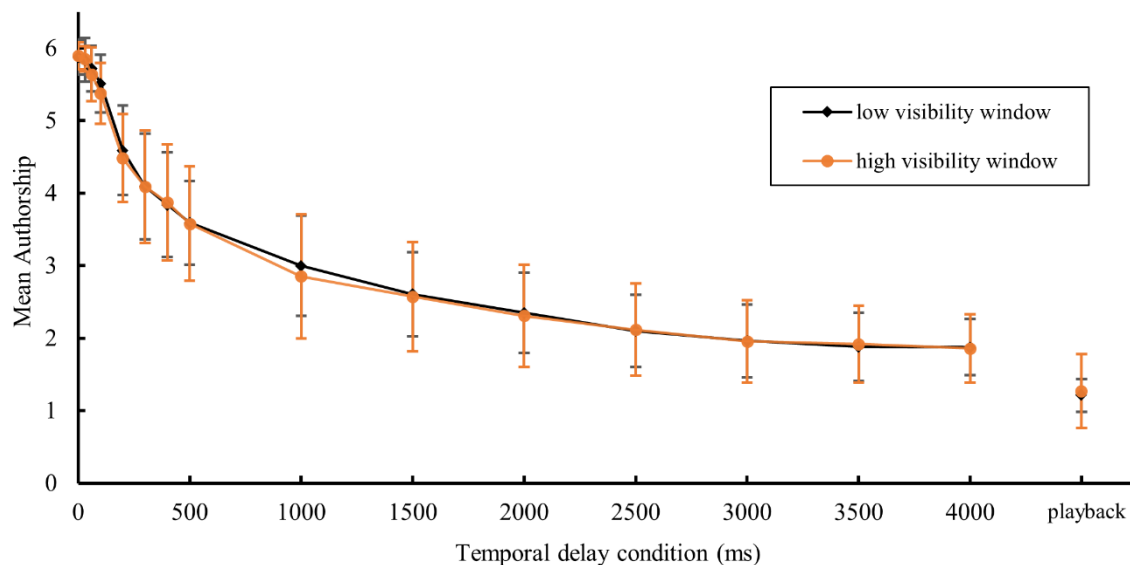


Figure 1.4. The authorship rating was measured on a 6-point scale to assess the degree of authorship, ranging from 1 ("I could not manipulate the window at all") to 6 ("I completely manipulated the window"). Mean authorship ratings are reported according to window visibility and delay, with means and standard deviations calculated across participants.

Reproduced from Kim et al. [84].

With a visibility $\times$ delay repeated-measures ANOVA, a highly significant main effect was revealed for delay ($F [2.199, 15.391] = 149.930$, $p < 0.001$, $\eta_p^2 = .955$), while neither

visibility ($F [1, 7] = .118, p = 0.741, \eta_p^2 = .017$) nor the interaction between visibility and delay ($F [2.874, 20.120] = .518, p = 0.667, \eta_p^2 = .069$) was significant. In line with the objective of identifying the critical temporal delay, post-hoc pairwise comparisons with Bonferroni correction showed that the mean authorship rating was significantly different from the 0 ms delay ($M = 5.883$) when the delay reached 200 ms ($M = 4.534, p = 0.031$) and beyond ($p < 0.05$).

A paired t-test comparing participants' mean authorship ratings between the playback condition and the 4,000 ms delay condition, was conducted to examine the decreased rating in the playback condition. The average authorship rating was greater in the 4,000 ms delayed window ($M = 1.87$) compared to the playback condition ($M = 1.24$). The mean difference of 0.63 was significant ($t [15] = 4.722, p < 0.001, \text{Cohen's } d = 1.465$), confirming that the mean authorship rating was lowest in the playback condition.

Behaviors

Response Time

The response time was evaluated to capture changes in behavior due to delays and window conditions. The means and standard deviations of the response times were calculated for participants' mean response times across each condition, considering the presence or absence of the target and window visibility (Figure 1.5).

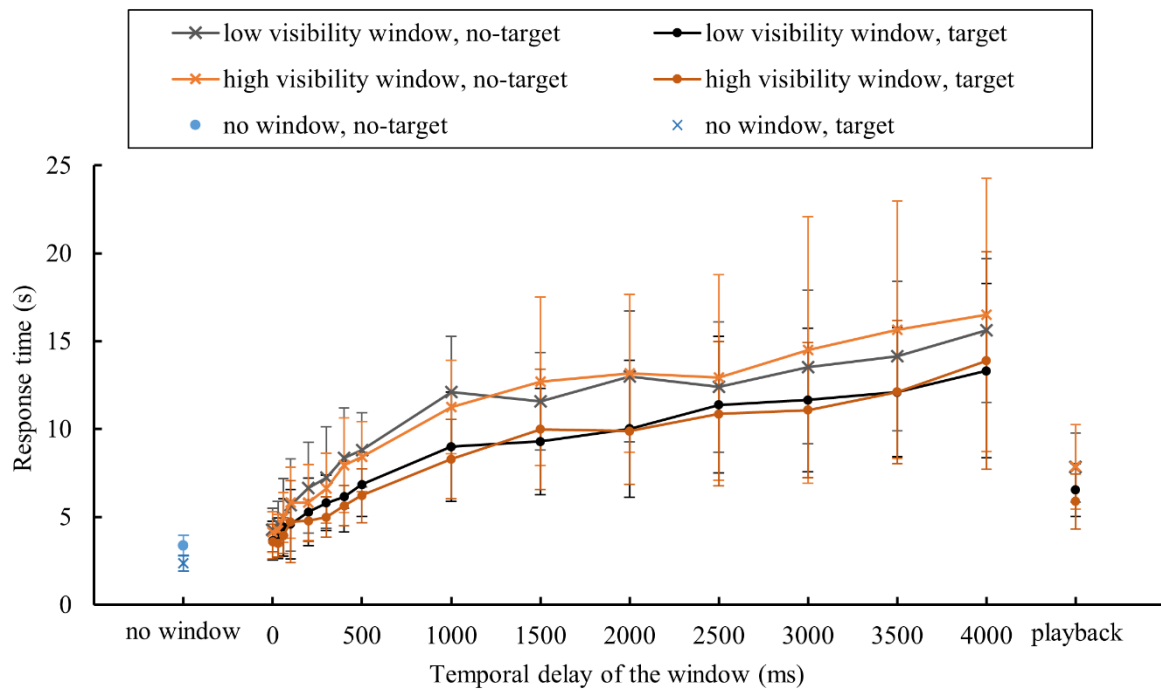


Figure 1.5. Means and standard deviations were displayed for the participants' response times. Reproduced from Kim et al. [84].

The visibility $\times$ target $\times$ delay repeated-measures ANOVA revealed a significant main effect of delay ($F [1.350, 9.450] = 53.906, p < 0.001, \eta_p^2 = .885$), indicating that response time increased as the delay increased. The main effect of target presence was also significant ($F [1, 7] = 18.180, p = 0.004, \eta_p^2 = .722$), with increased response times for stimuli without target ($M = 9586$ ms) compared to shorter response times for target stimuli ($M = 7693$ ms). The mean difference of 1894 ms was significant in the post-hoc pairwise comparison with Bonferroni correction ($p = 0.004$). Visibility did not show a significant effect ($F [1, 7] = .007, p = 0.936, \eta_p^2 = .001$). The interaction between delay and target presence was significant ($F [3.532, 24.723] = 4.265, p = 0.011, \eta_p^2 = .379$).

The effect of the gaze-contingent window was examined using a target $\times$ window presence repeated-measures ANOVA. Since the effect of the target has already been discussed, only the effect of window presence is reported here. The presence of the window

had a significant main effect ($F [1, 7] = 9.950$, $p = 0.016$, $\eta_p^2 = .587$). In the post-hoc pairwise comparisons adjusted using the Bonferroni correction, response times were significantly longer with the window present ($M = 3,952$ ms) compared to when the window was absent ($M = 2,869$ ms), with a mean difference of 1,083 ms ($p = 0.016$). The ANOVA confirmed that response times increased when using the gaze-contingent display compared to conditions without the window.

Regarding the current aim to identify the critical temporal delay, the post-hoc pairwise comparisons adjusted using the Bonferroni correction revealed that response time was significantly longer compared to the 0 ms delay ($M = 3909$ ms) when the delay reached 300 ms ($M = 6163$ ms, $p = 0.009$) and beyond ($p < .05$).

Eye Movements

The mean relative fixation duration frequencies were calculated across participants (Figure 1.6). In the no-window condition, the mode of the distribution was located around 125–175 ms, reflecting gaze behavior typical of a visual search [85]. With the gaze-contingent window, the mode shifted to 175–225 ms in the 0 and 30 ms delay conditions. As the delay increased, the mode continued to shift and the distribution exhibited negative skewness. Notably, from 200 ms to 500 ms delays, two distinct modes emerged: one around 175–225 ms and another that shifted further with increasing delay. For delays between 1,000 ms and 4,000 ms, including the playback condition, the mode consistently appeared at 150–225 ms. No significant differences were observed between the two window visibility conditions.

The average frequencies of relative saccade amplitudes were computed (Figure 1.7). Across all conditions, two primary distributions were observed: one mode located below 1° , and another around $7\text{--}8^\circ$. The minimum distance between the two characters was 8° , which probably contributed to the peak observed at $7\text{--}8^\circ$. The relative frequency of saccades under

1.5° increased with the gaze-contingent window compared to the no-window condition, and this frequency further rose as the delay increased. However, this increasing trend in relative frequency below 1.5° was inconsistent in the 1,000 ms to 4,000 ms delay range under the high visibility condition. Aside from this, no significant differences were observed between the two window visibility conditions.

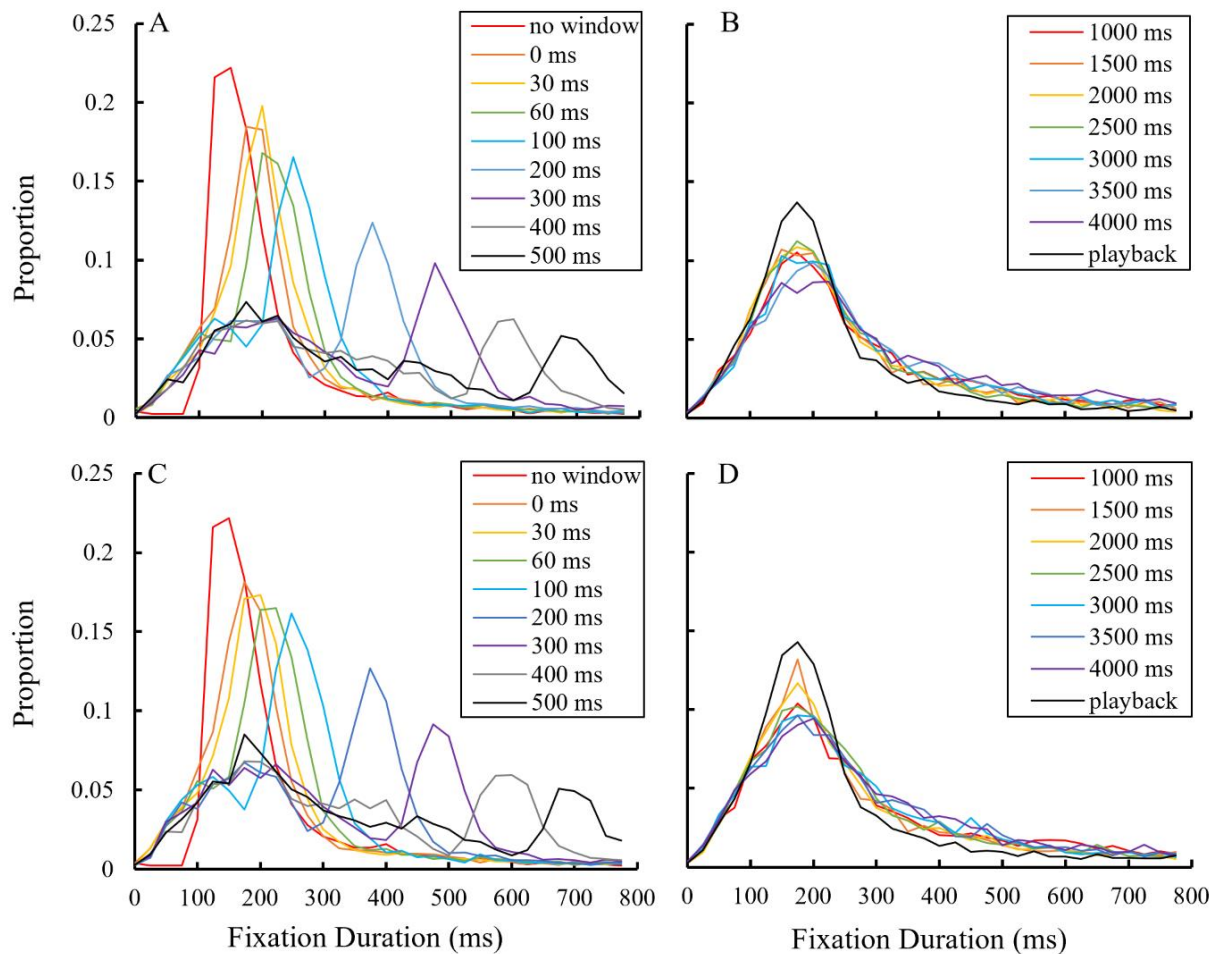


Figure 1.6. Relative distribution of fixation duration. For each participant, the frequency of fixation durations within each 25 ms interval across a 0–800 ms range was computed. The graph displays the averages across participants, with each bin's representative value positioned at the minimum value of that bin (e.g., the 25–50 ms bin is plotted at 25 ms on the x-axis). (A, B) distributions for the low visibility window. (C, D) distributions for the high visibility window. Reproduced from Kim et al. [84].

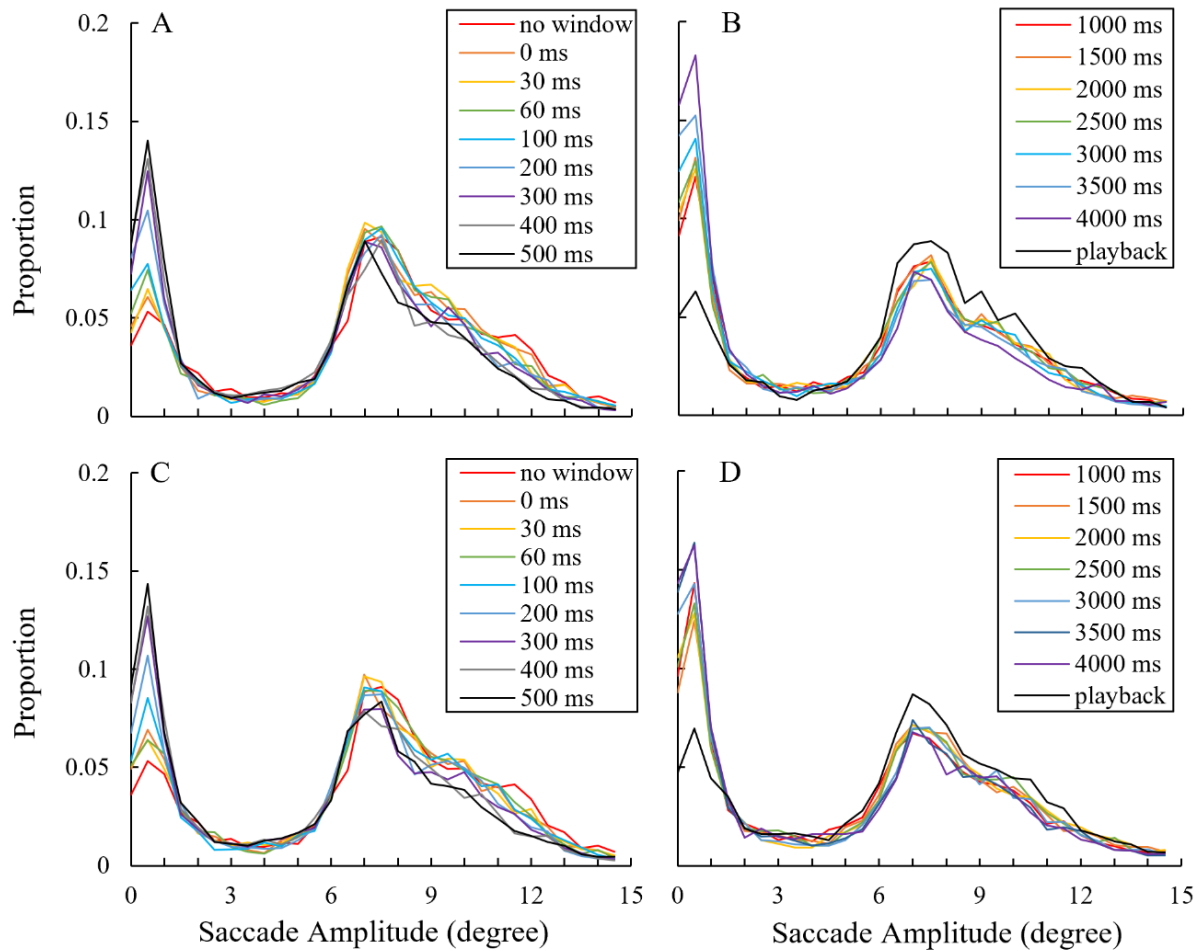


Figure 1.7. Relative distribution of saccade amplitude. For each participant, the frequency of saccade amplitudes within each 0.5° interval across the $0\text{--}15^\circ$ range was computed. The graph shows the average values across participants, with each bin's representative value positioned at the minimum value of that bin (e.g., the $1.5\text{--}2^\circ$ bin is plotted at 1.5° on the x-axis). (A, B) distributions for the low visibility window. (C, D) distributions for the high visibility window. Reproduced from Kim et al. [84].

Contrast Analysis

Following the observation of results within the 0-500 ms range, the location of the shifting mode in fixation duration as well as the cumulative distribution for saccades at 1.5 degrees was calculated. Additionally, the amount of change in each value was computed,

encompassing the rate of "Self" responses. Figure 1.8 presents these findings. Since visibility did not show any significant effects in either of the multivariate ANOVA analyses, the visibility conditions were combined for the purpose of plotting. It should be noted that the proportion of "Self" responses across the 0-500 ms range was pooled from Figure 1.3.

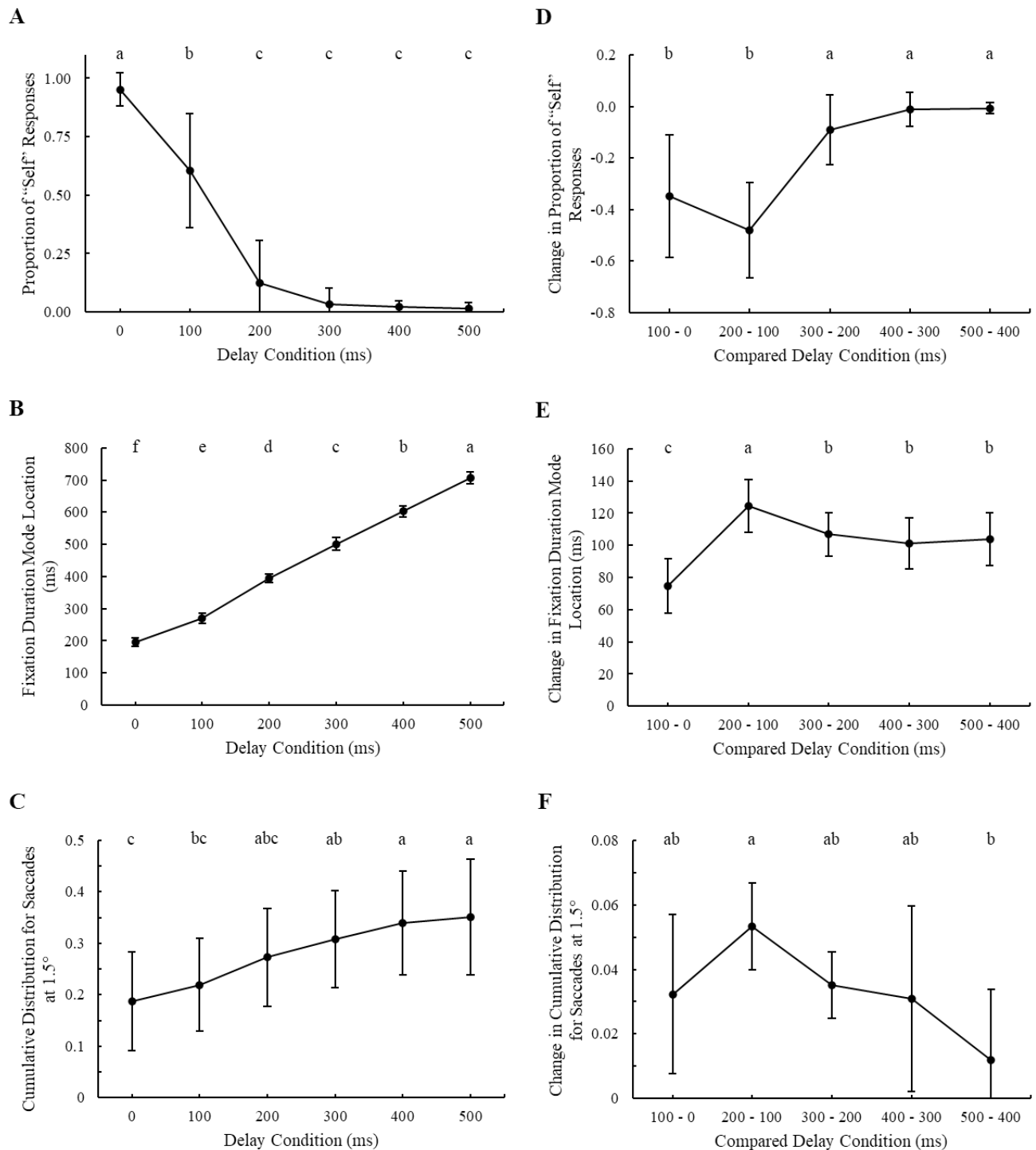


Figure 1.8. "Self" response proportions (A) has been re-plotted from Figure 1.3 with the

visibility conditions combined. Also displayed are the shifting mode location in fixation duration (B) and the cumulative distribution for saccades at 1.5 degrees (C). The amount of change was determined for each 100 ms delay increment for "Self" responses (D), fixation duration mode location (E), and cumulative distribution for saccades at 1.5 degrees (F). Visibility conditions were merged for plotting, as the variable did not show a significant effect. CLD graphs present the results of pairwise comparisons. Different letters denote statistically significant differences between delay conditions ($p < 0.05$), while identical letters indicate no significant difference between conditions. replotted from Kim et al. [84].

Two multivariate ANOVA analyses were conducted: one for delay $\times$ visibility and another for delay increase $\times$ visibility. In the delay $\times$ visibility MANOVA, a significant main effect of delay was found (Pillai's Trace = 1.659, $F[15, 252] = 20.781$, $p < 0.001$, $\eta_p^2 = .553$). No significant main effect was observed for visibility (Pillai's Trace = .037, $F[3, 82] = 1.036$, $p = 0.381$, $\eta_p^2 = .037$). The visibility and delay interaction was revealed as not significant (Pillai's Trace = .081, $F[15, 252] = .465$, $p = 0.956$, $\eta_p^2 = .027$). In the delay increase $\times$ visibility MANOVA, a significant main effect of the delay interval was found (Pillai's Trace = 1.254, $F[12, 210] = 12.571$, $p < 0.001$, $\eta_p^2 = .418$). No significant main effect was observed for visibility (Pillai's Trace = .006, $F[3, 68] = .145$, $p = 0.933$, $\eta_p^2 = .006$). The interaction between the delay interval and visibility was not significant (Pillai's Trace = .191, $F[12, 210] = 1.188$, $p = .293$, $\eta_p^2 = .064$).

Pairwise comparisons were conducted for both MANOVA analyses. The findings revealed that the proportion of "Self" responses decreased, while both the fixation duration mode location and the cumulative distribution for saccades at 1.5 degrees increased as the

delay lengthened. The change in "Self" responses between the 100-0 ms and 200-100 ms intervals was significantly different from the other three comparisons. However, the changes between the 100-0 ms ($M = -.348$) and 200-100 ms ($M = -.481$) intervals did not show significant differences from one another. Similarly, no significant differences were observed among the changes in the 300-200 ms ($M = -.090$), 400-300 ms ($M = -.010$), and 500-400 ms ($M = -.006$) intervals. These results suggest that the proportion of "Self" responses declined from 0 to 200 ms, then stabilized at delays beyond 300 ms. Regarding the fixation duration mode location, the amount of change in the 200-100 ms ($M = 124$ ms) and 100-0 ms ($M = 75$ ms) conditions was significantly different from the other three intervals. Additionally, a significant difference of 50 ms ($p < 0.001$) was found between the 100-0 ms and 200-100 ms conditions. The changes in the 500-400 ms ($M = 104$ ms), 400-300 ms ($M = 101$ ms), and 300-200 ms ($M = 107$ ms) intervals did not differ significantly from one another, indicating convergence at delays beyond 300 ms. For the cumulative distribution of saccades at 1.5 degrees, there was only one significant difference ($p = .003$) between the intervals of 500-400 ms ($M = .012$) and 200-100 ms ($M = .053$).

Discussion

This study explored three primary questions related to the relationships among temporal delays in action-feedback, the sense of agency, and eye movement. Participants' judgment of event authorship was assessed using a categorical question with three choices: self, delay, and other. The variation in the intensity of subjective authorship was measured with a rating question. Additionally, the impact of temporal discrepancies on participants' behaviors was examined. A secondary objective involved determining the critical temporal delay at which participants' experiences started to be influenced.

Sense of Agency

It was hypothesized that the findings would reflect those seen in limb modalities: 1)

participants would continue to attribute authorship to themselves rather than the computer, even with temporal delays in eye-gaze HCI, and 2) the perceived degree of authorship ("I") would diminish as temporal delays increased. This section discusses these two hypotheses.

When participants could control the eye-gaze HCI despite the delay, they did not attribute the computer as the event author. Categorical questionnaire responses showed that participants consistently distinguished between delayed and uncontrollable conditions (i.e., playback), as they reported "other" responses exclusively during playback trials. According to Synofzik et al.'s [26] framework, the contingency between motor intentions and outcomes is enough to identify the event author as oneself. Participants were instructed that their task was to visually search for a target character, likely planning a serial search strategy, as indicated by the longer response times when no target was present—though not twice as long as when the target was found. Nonetheless, during the playback condition, the window moved independently of participants' eye movements, forcing them to abandon their planned saccades and follow the window's unpredictable movement. Because trial conditions were randomly sequenced, participants couldn't anticipate whether the window would move independently of their actions. This mismatch between their saccade plans and the window's movement in the playback condition likely explains the "other" responses, as it reflects an incongruence between motor intentions and outcomes. This result aligns with previous research, including Sato and Yasuda's [18] first experiment, which demonstrated a weakened sense of agency when predicted outcomes did not match actual outcomes.

Moreover, participants were informed that the computer could control the window, a contextual cue that plays a critical role in authoring judgments, as specified by Synofzik et al.'s [26] model. This clear distinction in event author inference is also reflected in the rating questionnaire, which recorded the lowest scores in the playback condition. The continuous rating scale measured the intensity of influence exerted by the "self." When participants

attributed the event to the computer during the uncontrollable playback conditions, they chose the lowest rating of 1, indicating minimal "self" involvement in the outcome. The dissociation between response time and rating scores in the playback condition suggests that participants did not base their ratings on task complexity. This lowest rating is consistent with findings from social gaze experiments, where participants reported the least control over a virtual avatar that did not follow their gaze [29]. These results suggest that the experience of authoring in eye-gaze HCI shares similar characteristics with the sense of agency observed in social gaze contexts.

When a delay was present in the gaze-contingent window, participants typically responded with a "delay" response and did not attribute the event to another author (i.e., the computer). Previous studies have indicated that the sense of agency is maintained as long as the temporal action-effect mismatch does not exceed an internal time window of around 200-300 ms [34]. This time window is thought to buffer one's own movements. However, it seems unlikely that participants' action-effect time window could buffer their own eye movements for up to 4 seconds. So, how did participants still identify themselves as the event author under such extended delays? I propose that two aspects of the experimental action-feedback paradigm played crucial roles in shaping this judgment: the number of inputs (continuous versus single action-effect paradigm) and action-effect degree-of-freedom (directing versus a button push).

While "other" responses were not reported under delayed conditions in both this study and Farrer et al. [19], Farrer et al. [20] did find "other" responses with delays of around 700 ms. In my study, participants were not restricted in their response time, allowing them to continue each trial until they chose to terminate it. Indeed, when the delay exceeded 1000 ms, participants often extended the trial for 10 to 20 seconds, during which they could continuously navigate their gaze before forming their judgment of agency. In Farrer et al.

[19], participants similarly had the freedom to direct a joystick to update the virtual hand's position for 5 seconds. In contrast, in Farrer et al. [20], participants were limited to a single button press, and the feedback updated only once, with a grey ball appearing. The amount of sensory information gathered would be significantly different between continuously directing and pressing a button just once. This difference in action-effect paradigms likely explains the difference in the delay at which participants reported a "delay" response. In both my study and Farrer et al. [19], "delay" responses became dominant starting from 200 ms. Additionally, when comparing the inflection points reported in both studies, they both appeared at around 100 ms. Nevertheless, in Farrer et al. [20], the inflection point was observed at approximately 330 ms.

It is speculated that the decrease in "self" response proportions can be understood through the internal comparison process described by the comparator model. The current contrast analysis (Figure 1.8) revealed significant drops in "self" response proportions from 0 to 200 ms, with proportions converging to nearly 0% by 300 ms. Notably, at a 200 ms delay, the "self" response proportions dropped to around 25%. Here, I propose that participants consistently perceived delays beyond 200-300 ms as exceeding the time window associated with the sense of agency (referred to as body agency by Wen [34]), which is typically around 200 ms. According to Synofzik et al.'s [26] framework, this perception can be attributed to the incongruence between the expected state (a high-resolution image) at the attended object and the actual state (a blurred image). Despite this, the delay did not disrupt participants enough to cause them to abandon their planned saccades (as discussed in detail in the Behavior section regarding their eye movement strategies). As a result, this judgment led participants to report "delay" responses rather than "other" responses. Synofzik et al. [26] also discussed the phenomenon of "self-despite-mismatch," where individuals continue to attribute actions to themselves even when there is a significant mismatch between predicted

and actual sensory outcomes.

The GLMM analysis of the categorical responses revealed that window visibility significantly influenced participants to perceive delays at relatively shorter durations. In the high visibility condition, where the background brightness was reduced compared to the window, more salient features were provided—such as motion perception, brightness contrast, and differences in spatial resolution between the inside and outside of the window—compared to the low visibility condition, which relied solely on spatial resolution. These enhanced features might have been processed more quickly, leading to earlier detection of the delay. However, this early detection in the high visibility condition was not significant in the current contrast analysis or in the rating questionnaire. The contrast analysis MANOVA took into account not only the “self” response but also eye movement behaviors, which might have reduced the sensitivity required to detect a significant visibility effect. Additionally, the rating questionnaire may have struggled to capture subtle differences within the 0 to 200 ms delay range, as the rating scores did not converge to specific values for delays beyond 300 ms. The categorical response, however, showed convergence, producing similar outcomes across visibility conditions at delays over 300 ms, which could explain the higher sensitivity to significant statistical results within the 0 to 200 ms delay range. The small sample size in this study could have further contributed to the lack of statistical power. These nuances suggest that the difference in spatial resolution inside and outside the window was enough to be noticed when detecting the delay, rendering visibility a minor effect. Nevertheless, these findings offer valuable insights for designing visual elements in HCI applications.

While the “delay” response became dominant at delays beyond 200-300 ms, indicating convergence in the categorical response results, the mean rating score continued to decrease after a 300 ms delay. Similar decreases in continuous scales have been consistently observed in numerous previous studies [18,21–24], suggesting that participants’ ratings

reflect a general phenomenon across different modalities. The score continued to decline even beyond the point where "delay" responses had converged in the categorical reports, eventually reaching convergence as the delay approached 4000 ms.

An interesting observation from the rating scores is that the graph showed a concave upward pattern as the delay increased, meaning that the reduction in scores for the same incremental delay gradually lessened as the delay grew longer. This suggests that participants' sensitivity to judging their own influence on the outcome diminished with longer delays. One possible explanation is that the internal comparison process described by the comparator model contributed to high sensitivity at shorter delays. The decrease in ratings from 0 to 300 ms was relatively steep, compared to the more gradual decline in scores afterward. Although this concave upward trend was observed in the current study, it has not been explicitly confirmed in previous research with limb modalities [18,21–24], limiting further discussion. This indicates a need for additional research to replicate these findings in different modalities and conditions. Nevertheless, the mean rating scores provide insight into how participants judge their influence in authoring the HCI over longer durations, offering information that categorical reports alone could not capture.

Behaviors

This study revealed that eye movements were altered with the introduction of delays. Initially, it was hypothesized that different task strategies would be adopted by participants in response to these delays. However, specific predictions about the changes in eye movements were not made, making this investigation exploratory. In this discussion, the task strategies employed by participants as delays varied are explained based on the observed eye movement data. Additionally, the contrast analysis is examined to explore how these changes in eye movements might be connected to the sense of agency.

First, two key points related to the strategies observed in the fixation duration

distribution (Figure 1.6) are discussed: 1) the shifting tendency of the mode in delay conditions ranging from 0 to 500 ms, and 2) the unimodal distribution observed in delay conditions of 1,000 ms and beyond. Two distinct fixation strategies are proposed to account for each point. The first strategy, referred to as the "wait-and-read strategy," can be explained by considering a scenario where participants make a saccade to an arbitrary character on the array during the visual search. After making the saccade, participants keep their gaze fixed on that location until the gaze-contingent window updates to that position, allowing them to then observe the Chinese character. This waiting period for the window update explains the mode shifting observed in the distribution for delays under 500 ms, with the latter mode occurring between 200 and 500 ms (Figure 1.6). The second strategy, termed "pre-registering," offers a different approach. In this scenario, after making a saccade to an arbitrary character during the visual search, participants do not remain at the fixed location while waiting for the window to refresh. Rather, they proceed to make further saccades to other characters. When the delay is sufficiently long, participants repeat these saccades until they perceive that the window has begun to update in line with the first and subsequent saccades. Once the window starts updating, it tracks the sequence of fixations that participants had pre-registered with their gaze since the first saccade. At this point, participants follow the window to inspect the character within it. The primary difference between these two strategies lies in their approach to handling the delay: in the "wait-and-read" strategy, participants maintain their gaze at the fixated location until the window updates, while using the strategy of "pre-registering", participants continue to make saccades to new locations without waiting for the window to update.

The assumption about the "pre-registering" strategy was informed by two key sources: direct observations made during the experiments and a review of the experimental sessions replayed using the EYELINK Data Viewer's playback animation feature. This

strategy was identified through repeated eye movements following the same trajectory. For example, participants moved their gaze in a clockwise or counterclockwise pattern during delays of 1,000 ms and above. For delays shorter than 1,000 ms, they frequently alternated saccades between two characters. The reduced upward slope of the response time function at delays over 1,000 ms (Figure 1.5) supports the notion of a strategy change. It is assumed that this “pre-registering” strategy manifested as a distribution with a mode located at approximately 200 ms of fixation duration, as observed in Figure 1.6 under the following delay conditions: 1) the earlier mode from the 200 to 500 ms delay range, and 2) the mode from the 1,000 to 4,000 ms delay range. The 200 ms likely reflects the time required to memorize an arbitrary Chinese character and its location, corresponding to the mode observed for the non-delayed window (“0 ms” in Figure 1.6). While the “pre-registering” strategy appeared exclusively at temporal delays of 1,000 ms and above, it is possible that both the “wait-and-read” and “pre-registering” strategies occurred simultaneously in the 200 to 500 ms delay range, resulting in the observed bimodal distribution.

From these considerations, two key discussions arise: 1) the transition from the “wait and read” strategy to the “pre-registering” strategy as temporal delays increased, and 2) the shifting mode observed in fixation durations.

First, it appears that participants adopted the strategy that allowed them to view all Chinese characters in the shortest possible time. The use of the phonological loop as a verbal encoding strategy during the visual search might have been a factor. Although articulatory suppression was not employed to mitigate this strategy, participants were asked to verbally report the characters they could read after the experiment. They reported recognizing only 1–3 characters out of the 239 presented. This suggests that the phonological loop strategy was likely not used during the visual search. As participants reached the temporal limits of their sensory memory, they likely needed to abandon the “wait and read” strategy at extended

delays, ultimately relying solely on the "pre-registering" strategy for delays longer than 1,000 ms. This shift in strategy can be understood without invoking the framework of the sense of agency. The downward trend in the rating responses beyond 1,000 ms of delay (Figure 1.4) indicates that participants' sensitivity to evaluating their own influence on the outcomes diminished with delays over 1,000 ms. These findings suggest that the shift to the "pre-registering" strategy was driven by memory limitations rather than changes in the sense of agency.

Second, the mode shift. The data confirmed that the proportion of "self" responses decreased significantly in the 0 to 500 ms delay range. To explore the relationship between this mode shift and the sense of agency, I focused on how the mode shift was affected by the delay within this range. The contrast analysis showed that the mode shift was most pronounced between 100 and 200 ms, a period that also saw a significant drop in "self" responses, which fell to about 25% at 200 ms. This suggests that participants' awareness of the delay may have contributed to the significant shift in the mode. The comparator model might offer an explanation, as bodily agency is understood to operate within a 200 ms time window [34]. The "wait-and-read" strategy implies that the mode should shift by 100 ms for every 100 ms increase in delay, a pattern observed in the 300 to 500 ms range. Interestingly, the mode shift from 0 to 100 ms was shorter than those in the 300 to 500 ms range, whereas the 100 to 200 ms shift was longer. This shorter mode shift from 0 to 100 ms likely contributed to the more substantial shift from 100 to 200 ms. The delay range from 0 to 100 ms corresponded to a significant drop in "self" responses to 50%. According to the comparator model, it is plausible that participants' action refinement was lessened when bodily agency was maintained (as the delay did not exceed the time window), and then intensified when the mismatch was detected at 200 ms.

When analyzing the saccade amplitude distribution, an increased relative frequency of

0–1.5 degree saccades was observed as temporal delay increased (Figure 1.7). Although the capability of camera-based eye-tracking to accurately capture microsaccades is questionable, insights can be drawn from previous research on microsaccades to infer potential causes. Privitera et al. [86] discovered that microsaccade frequency rose during visual searches with ambiguous targets. Additionally, microsaccades help in obtaining spatial detail [87] and repositioning gaze during tasks that require fine spatial discrimination [88]. Another study demonstrated that participants could voluntarily generate microsaccades when re-fixating on a memorized target [89]. Microsaccades are also considered indicators of covert attention allocation [90,91] and tend to become less frequent as task difficulty increases in both non-visual [92] and visual tasks [93].

While the Chinese characters used in this experiment required participants to discern fine spatial details, particularly when discriminating between characters with similar partial shapes, the increasing trend in microsaccades with delays is challenging to explain through task difficulty alone. Previous studies have shown a decrease in microsaccade frequency with increased task difficulty, making this an unlikely explanation. Instead, a detachment between eye movements and attention allocation may be a more plausible interpretation. During the visual search, participants might shift their attention to an arbitrary character, but due to the temporal delay in the gaze-contingent window, the character appears blurred when fixated on. This delay creates a spatial gap between where participants are looking and where the window is positioned. Participants might then allocate their attention to the window while their gaze remains on the intended location. With increasing delay, and consequently, the spatial distance between the window and eye movements grows, the need to allocate covert attention to the window might also intensify, leading to the observed increase in microsaccades.

The increase in saccades under 1.5 degrees could be related to the sense of agency

through the role of covert attention. The contrast analysis was conducted on the cumulative distribution at 1.5 degrees, and although the result was not statistically significant, the increase was most pronounced in the delay range from 100 to 200 ms. This finding suggests that participants' covert attention distribution might be heightened during this delay range as they perceive the delay. I propose that covert attention distribution might play a role in the internal mismatch comparator, as attention is crucial for comparing an action with its effect [94]. Taylor [95] further proposed that corollary discharge (or efference copy in the context of eye movements) is generated from an attention-shifting signal rather than a motor signal and suggested that our consciousness draws internal cues from this signal. Thus, the current findings hint at a close relationship between the distribution of attention and the sense of agency. This suggests that as participants perceive a delay, their covert attention distribution becomes more prominent, potentially contributing to their internal sense of agency by enhancing the comparison between intended actions and observed outcomes.

In delays exceeding 1,000 ms, where participants adopt the "pre-registering" strategy, their covert attention may be directed toward the anticipated position of the window, ahead of its current location, based on their memory of the preregistered eye movement trajectory, while simultaneously reading the character at their current gaze location. This interpretation aligns with previous studies that suggest microsaccades are often generated when re-fixating on a memorized target [89]. The simultaneous deployment of covert attention toward the delayed window and along the anticipated trajectory might explain the relatively high frequency of 0–1.5 degree saccades observed under these conditions.

Critical Temporal Delay

As a secondary objective, the critical temporal delay at which user experience was interrupted was examined. Post-hoc analysis identified this critical delay as 300 ms for response time, 200 ms (mean authorship rating), and same 200 ms for the categorical

judgment. The critical delay of 200 ms aligns with the current contrast analysis results, where significant changes were observed in both the categorical authorship report and fixation duration. As speculated in the discussion of the contrast analysis, this may indicate the time window associated with the feeling of agency [34], suggesting a connection between the alteration of user experience and bodily agency.

However, the current results revealed an inconsistency between subjective reports and response time. Although different measures may have varying intrinsic sensitivities to detecting delays, this inconsistency suggests a possible power deficit in the analysis, which included only eight participants. While this limitation does not conclusively rule out the possibility of influences at shorter delays, it does indicate that further investigation is needed. An inspection of the proportion of categorical authorship responses, fixation duration distribution, and saccade amplitude distribution revealed a tendency to deviate starting at the 30 ms delay, hinting at an early effect on performance. Future research with a larger participant pool may uncover a critical delay shorter than the 50–70 ms range suggested by earlier studies that used a 3D graphic scene with a head-mounted display [68] and a natural scene displayed on a 2D monitor [69].

An intriguing observation was that when the temporal delay extended beyond the critical threshold, users' performance was noticeably disrupted; however, they were still able to manipulate the gaze-contingent window and complete the task, even at a delay of 4,000 ms. This is similar to the satellite teleoperation scenario, where participants were able to control a robotic arm with their hands despite a 7-second temporal delay [49]. Nonetheless, further research with a different experimental design is necessary to fully understand users' delay tolerance in task completion, especially since the current experiment did not impose a time limit on the task. Additionally, manipulating the number of Chinese characters in the stimuli could offer another dimension for exploring how delay impacts task performance.

Gaze-Contingent Display

In this section, the stimuli used in the experiment is evaluated. A 3×3 virtual grid of Chinese characters with a gaze-contingent window was implemented to address the study's main questions while mitigating the decrease in saccade amplitude associated with peripheral blur, as indicated by previous research [76,78–80,82]. The results showed that the gaze-contingent window increased response time and influenced gaze behavior. Specifically, fixation durations were longer with the window, where the mode of distribution shifted from 125–175 ms (without window) to 175–225 ms (with window). Regarding the saccade amplitude, there was a higher frequency of saccades under 1.5° with the window. However, saccade amplitude did not decrease overall, as $7\text{--}8^\circ$ saccades consistently appeared across all conditions, likely due to the 3×3 virtual grid. In sum, while the gaze-contingent window did affect participant behavior, it did not hinder the ability to address the study's main questions.

Conclusion

Chapter 2 demonstrated that temporally delayed gaze-contingent displays significantly impact the sense of agency and eye movements. By expanding the scope of previous research on agency involving limb and hand movements to include eye movements, we revealed the interplay between delay perception and action refinement. Contrast analysis of the subjective reports and eye movements suggests that the previously suggested narrow time window (200 ms) and action refinement in the comparator model [34] would also explain the current results. The systematic investigation into how delay influences task strategies and action refinement in eye-gaze HCI had barely been revealed in hands or limbs research. The current findings suggest that eye-gaze interfaces propose that the eye-gaze HCI system could detect changes in gaze behavior as indicators of delay perception, contributing for the system to capture the user's experience. Understanding the critical temporal delays that disrupt the sense of agency can guide the development of delay-tolerant interfaces,

enhancing user experience in first-person perspective-based applications such as remote surgery, teleoperation, and virtual reality.

Chapter 3

Mitigating the temporal gap between a user's eye movements and the visual feedback in an HCI system is a complex task, especially when the HCI suffers from inconsistent jittery delays in teleoperation scenarios. While inconsistent "jittery" delays in data transmission are known to significantly impact the user experience in video streaming [50,51,96], comprehensive perceptual studies in teleoperation settings are scarce. The unpredictable nature of RTTs complicates the creation of thorough models, as variables like communication locations, type and size of data, and a wide range of perceptual tasks can easily alter RTTs. Although various delay dynamics models across regions [52–56] have demonstrated that RTTs typically follow a highly right-skewed distribution, empirical studies on user performance [45,49,57,97] and cognitive workload [98–102] have not systematically examined RTT variations. Additionally, due to the specific nature of these scenarios, generalizing findings to a perceptual context involving the sense of agency and temporal discrepancies is difficult. In considering how internal mismatches (e.g., delay) are processed in the comparator model, current experiment was particularly interested in understanding the impact of jitter on this process. Are the prediction errors we become aware of related to the minimum, maximum, or average value of the jittery distribution?

This study introduced action-effect jittery temporal delays and eye-gaze HCI to explore the sense of agency. Subjective reports, while limited by their susceptibility to biases and criticism for only reflecting the conceptual judgment of agency, hold value as they enable direct comparisons with prior research [19,20] and Experiment 1. Additionally, this study examined the impact on participants' eye movement behavior, proposing altered behavior patterns as indicators of a disrupted user experience. The current goal was to identify the statistical attributes within the jittery delay distribution that best explain the effect on the sense of agency and eye movements (e.g., the lowest, average, or highest values of the

distribution).

Overall, this study was driven by the following research question:

- In what ways does a jittery delay impact the sense of agency during eye movements?
- How does temporal discrepancy alter eye movement patterns?
- Which specific statistical factors are most effective in explaining the interplay between the sense of agency and eye movements when a jittery delay is present?

Materials and methods

Participants engaged in visual search tasks using a bi-resolution gaze-contingent display, which limited the high-resolution display area to a square shape in the central field of view (see Figure 2.2 for a schematic representation). Stimuli were presented in two distinct regions with a clear border: a 3x3 array of Chinese characters within the high-resolution window and a blurred array in the surrounding area. Participants were directed to locate a specific character, matching the one shown in the center, among the eight surrounding positions. The target character might be absent or appear just once in the stimuli. After each task, participants answered two questionnaires. Each participant completed 15 sessions (270 trials). Each session was composed of 18 trials: two practice trials, two series of seven delayed trials each, and two playback (i.e., uncontrollable) trials. In the trials of practice, the gaze-contingent window was updated without any intentional delay. In delayed trials, the window's update followed a truncated normal distribution delay between eye movement and display update. The playback trial assessed whether the questionnaires could capture participants' misattribution, replaying the participant's gaze behavior from one of the practice trials instead of mirroring their real-time gaze. Following the practice trials, the delayed and playback trials were presented in a randomized order.

I set σ to 50 ms and established seven conditions for μ : 0, 50, 100, 150, 200, 300, and

400 ms with a truncated normal distribution (from μ to $\mu+2\sigma$ where μ is the mean and σ is the standard deviation, Figure 2.1A) for the jittery delay. Temporal discrepancies were randomly generated for each eye movement data point. Occasionally, the gaze-contingent window displayed eye data in reverse order due to random generation properties (w3 and w4 in Figure 2.1B). A sample of the randomly generated temporal discrepancy throughout a trial is shown in Figure 2.1C as a time function and Figure 2.1D as a distribution ($\mu = 150$, $\sigma = 50$).

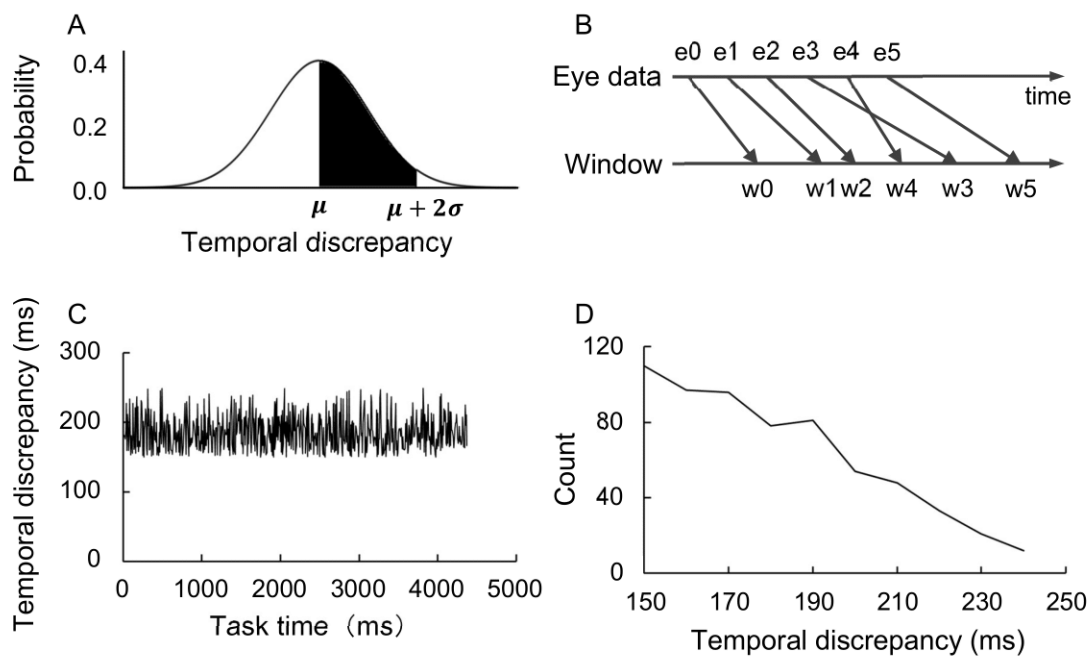


Figure 2.1. A, top-left: The temporal discrepancy between recording eye movements and updating the gaze-contingent display was randomly varied for each data point within a range from μ to $\mu+2\sigma$ (with a normal distribution). B, top-right: Different data points experienced varying delays, occasionally resulting in a reverse-order display. C, bottom-left: A sample trial demonstrating jittery delay as a function of time. D, bottom-right: The distribution of jittery delays ($\mu = 150$, $\sigma = 50$) with 10 ms bins, where each bin's representative value is placed at its minimum (e.g., the 240-250 ms bin value is shown at 240 ms). Reproduced from Kim et al. [103].

Previous studies have indicated that RTTs generally follow highly right-skewed distributions, such as gamma [52,53], exponential [54,55], or Pareto distributions [56]. The

current objective was to examine each statistical aspect of jittery delay (lowest, mean, and highest) and highlight the differences among these aspects. Although highly right-skewed distributions often have similar minimum and mean values, the truncated normal distribution offers defined boundaries, facilitating the differentiation between the statistical characteristics of jittery delay. Therefore, I adopted this distribution. The delay ranges were selected to include areas where subjective authorship reports indicated a reduction, while also providing a realistic representation of teleoperation conditions.

Participants reported the extent to which they felt their eye movements caused the window to move using a six-point scale, a method commonly used in previous research [18,21–24]. Furthermore, this study employed a questionnaire from Farrer et al. [19,20] to investigate the effect of action-effect temporal delay on authorship across three levels: perfect control, partial control, and control by an external agent

Participants also completed two sessions, each consisting of 36 trials, of a visual search task without a gaze-contingent display. In this condition, they did not answer any subjective questionnaires, aiming to assess the impact of the gaze-contingent window on eye movements and task difficulty. Each session presented 36 random images selected from a pool of 540 images.

Participants

Twelve individuals participated, but data from one were excluded due to technical issues with keyboard responses. Therefore, data from 11 participants (six undergraduates, three master's students, and two doctoral students from Tokyo Institute of Technology) were analyzed. Two participants did not conduct the no-window condition. The participants, aged 19-27 (average age: 23.9, SD: 2.4), included two women and nine men, with three Japanese, six Korean, and two Chinese participants. Only those who could read Chinese characters were recruited. Participants earned a 1,000-yen Amazon gift card per hour of participation.

All participants either had normal vision or wore corrective lenses and gave their informed consent. The study was conducted according to the Declaration of Helsinki guidelines and approved by the Tokyo Institute of Technology Ethics Review Committee for Epidemiological Research. Data were collected from April 26, 2022, to December 15, 2022.

Apparatus

Participants completed the experiment in a dark room using the same apparatus as in Experiment 1, with software modified from that used in Experiment 1.

Stimuli

The same stimuli and 540 images from the low-visibility condition in Experiment 1 were used (Figure 2.2). Participants did not adjust their gaze to view two characters simultaneously, an unintended behavior (details in Appendix 2.1).

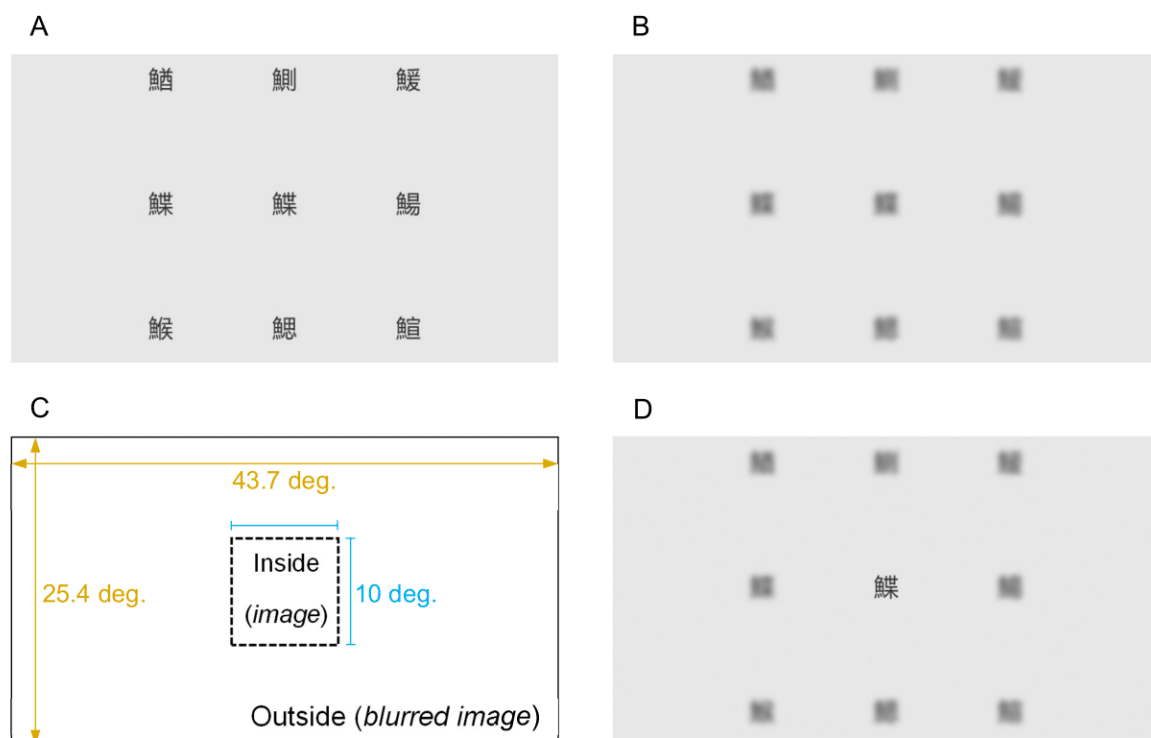


Figure 2.2. A, top-left: A stimulus featuring nine Chinese characters in a three-by-three virtual matrix, randomly generated. All nine characters had the same number of strokes and

included a fish radical. A total of 540 images were created, with the target character either appearing or not appearing in each image. B, top-right: One of the 540 images was randomly chosen and blurred. C, bottom-left: Within the gaze-contingent display (rectangle with solid borders), a virtual window following gaze (square with dashed borders) was centered on the participant's fixation location. The image was displayed within the window, with a blurred image surrounding the central area. D, bottom-right: An example stimulus when the participant focused on the central Chinese character. Reproduced from Kim et al. [103].

Procedure

Each trial started with a drift check, followed by a 650 ms display of blurred images, then the visual search began. This duration prevented participants from expecting deliberate temporal discrepancies. During the waiting period, participants were instructed to fixate on the monitor's center. The visual search ended when participants identified and reported the target's location and pressed enter. They then completed the authorship questionnaire to assess whether they perceived the window movements as aligning with their gaze (self), experiencing a temporal delay (delay), or being controlled by another agent (other). They rated the degree of their eye movements' influence on the window's movement on a six-point scale, where six meant “I fully manipulated the window” and one meant “I could not possibly manipulate the window.”

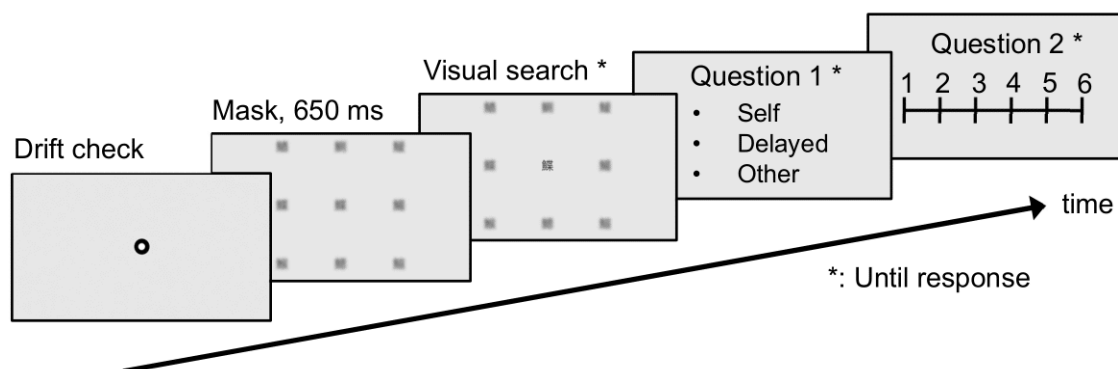


Figure 2.3. Trial sequence: After the drift check, participants awaited the appearance of the

gaze-contingent window, which allowed them to read a Chinese character only within the window. Once they responded, participants filled out two questionnaires. The drift check and questionnaires were enlarged for clarity. Reproduced from Kim et al. [103].

Analyses

All statistical analyses were conducted using the average values for each condition, representing each participant.

To investigate the delay's impact on the sense of agency, the current study's first question, I utilized two generalized linear mixed models (GLMM; SPSS 24.0) to analyze the μ value's effect on categorical authorship reports. In particular, for trials that involved delays—excluding playback and no-delay trials—I compared 'self' responses to 'delay' responses and 'other' responses to 'delay' responses. This comparison was made using a multinomial probability distribution and a generalized logit link function. The participant was treated as a random effect, using variance components as the covariance structure to account for unexplained variance, while the μ value was considered a fixed effect. Because of data loss, four responses were discarded, leaving 2,306 responses for analysis. For the mean authorship rating scores, a one-way ANOVA with repeated measures was performed on delayed trials. Sphericity was assessed using Mauchly's test, and degrees of freedom were adjusted based on Greenhouse-Geisser estimates.

To address the current second research question on the impact on participants' behavior, I first analyzed response times. Response time was evaluated as an indicator of task complexity. The time from stimulus onset to the participants' final input was recorded, with prior inputs being disregarded due to frequent repeated registrations when participants repositioned their hands on the keypad. I conducted a target $\times$ delay repeated-measures ANOVA on response time under the delayed conditions, adjusting degrees of freedom

similarly. For pairwise comparisons, Bonferroni correction was applied. For eye movements, fixation counts per trial and per second, relative distribution of fixation durations, and saccade amplitudes were examined, all concerning the μ value. Minimum thresholds were set to identify saccades (0.1° , $30.0^\circ/\text{s}$, and $8000.0^\circ/\text{s}^2$). Fixations that occurred outside the stimuli, started immediately before or after a blink, or lasted less than 120 ms were excluded from the current analyses. The relative distribution of fixation durations was calculated in 25 ms intervals within a 0-800 ms range and the distribution of saccade amplitudes in 0.5° intervals within a $0-15^\circ$ range. Due to variability in response times across trials and participants, I utilized relative distributions. For fixation counts per trial, a one-way repeated-measures ANOVA (delay) was performed under the delayed conditions. Sphericity was evaluated using Mauchly's test, and degrees of freedom were adjusted according to Greenhouse-Geisser estimates.

The third aim was to identify the statistical attributes (e.g., lowest, mean, or highest values from the distribution) that most significantly influence the sense of agency and eye movements. I compared the current results (categorical authorship reports and fixation count per second) with those of the previous experiments, which implemented a constant delay ranging from 0 to 4000 ms (Experiment 1). The previous study used the same gaze-contingent window and included no-window and playback conditions, similar to the current study. For comparison, I used data from delayed conditions ranging from 0 to 400 ms, retrieved from Experiment 1. Only the average (μ) value within the jittery distribution used in the current study was considered, representing the minimum value of that jittery distribution. If the comparison showed no significant difference between the two experiments, I concluded that the minimum value was the most influential attribute.

Using fixation count per second, I conducted a two-way mixed ANOVA for the five delayed trials (0, 100, 200, 300, and 400 ms), with delay as a within-subjects factor and type

of experiment (constant delays and jittery delays) as a between-subjects factor. For categorical authorship reports, I applied the least-squares method to fit a sigmoid function to the proportion of each of the two categories in the authorship report (“self” and “delay”). Other responses were omitted due to their rarity across both experiments. For both experiments, two regression curves were fitted for each participant, resulting in 38 curves. $Y = 1/(1 + e^{a(x-b)})$ (i.e., sigmoid function) was used, with a and b as adjustable parameters. The parameter b represents the delay at which the response proportion reaches the 50% threshold, while a indicates the slope of the threshold. A positive a represents an increasing slope, and a negative a indicates a decreasing slope. Excluding fits with an R-squared value less than 0.70 resulted in the removal of one delay response curve from the current jittery delay experiment, leaving 37 curves for comparison. The 50% threshold was selected as the threshold was reported to reliably represents shifts in the subjective sense of agency [20,104,105]. Steeper slopes indicated higher certainty in participants' responses. Consequently, I fitted a sigmoid function as described by Farrer et al. [20]. With the fitted values of slope (a) and threshold delay (b), I conducted a one-way multivariate ANOVA for the delayed trials, treating experiment type as a between-subjects factor.

Results

Subjective authorship

The categorical authorship report comprises three response levels (“self”, “delay”, and “other”). For each condition, I calculated the mean and standard deviation of the proportion of each response type, as shown in Figure 2.4. The GLMM results indicated a significant main effect of delay when comparing 'self' to 'delay' responses ($b = -0.018$, $SE = 0.002$, $t = -8.913$, $p = .000$), indicating fewer 'self' responses than 'delay' responses as μ increased. All parameters and coefficients for the GLMM are available in Appendix Table 2.3. When comparing 'other' responses to 'delay' responses, the main effect of delay was not significant

($b = 0.000$, $SE = 0.002$, $t = 0.122$, $p = .903$), with only the intercept showing significance ($b = -3.387$, $SE = 0.507$, $t = -6.683$, $p = .000$), indicating the infrequent choice of 'other' responses in delay conditions. All parameters and coefficients for the GLMM can be found in Appendix Table 2.4.

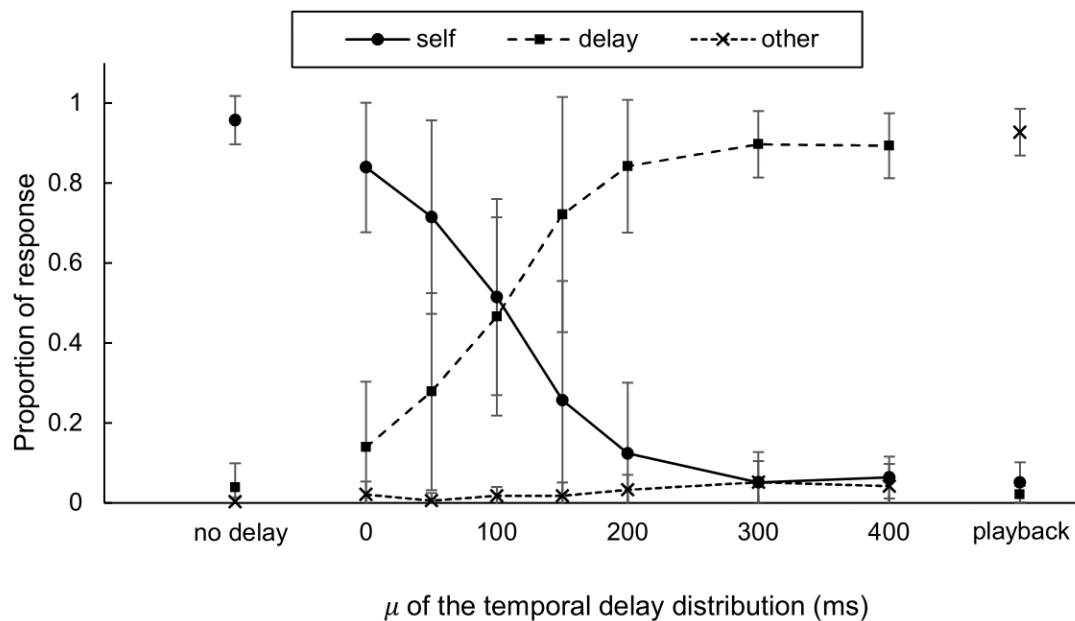


Figure 2.4. The proportion of each categorical response (self, delay, or other). Means and standard deviations were calculated across the 11 participants. Reproduced from Kim et al. [103].

Table 2.1 provides the means and standard deviations for each fitted value of slope (a) and threshold delay (b) across the various categorical authorship responses in both current experiments and prior Experiment 1. For a visual depiction of these results, refer to Figure 2.5. It is important to note that 'self_jitter' and 'delay_jitter' in Figure 2.5 represent the average proportions of 'self' and 'delay' responses shown in Figure 2.4. The multivariate ANOVA, which evaluated the influence of experiment type (constant vs. jittery), did not reveal a significant main effect ($F [4, 13] = .459$, $p = .764$, partial $\eta^2 = .124$). All parameters for the MANOVA are available in Appendix Table 2.5.

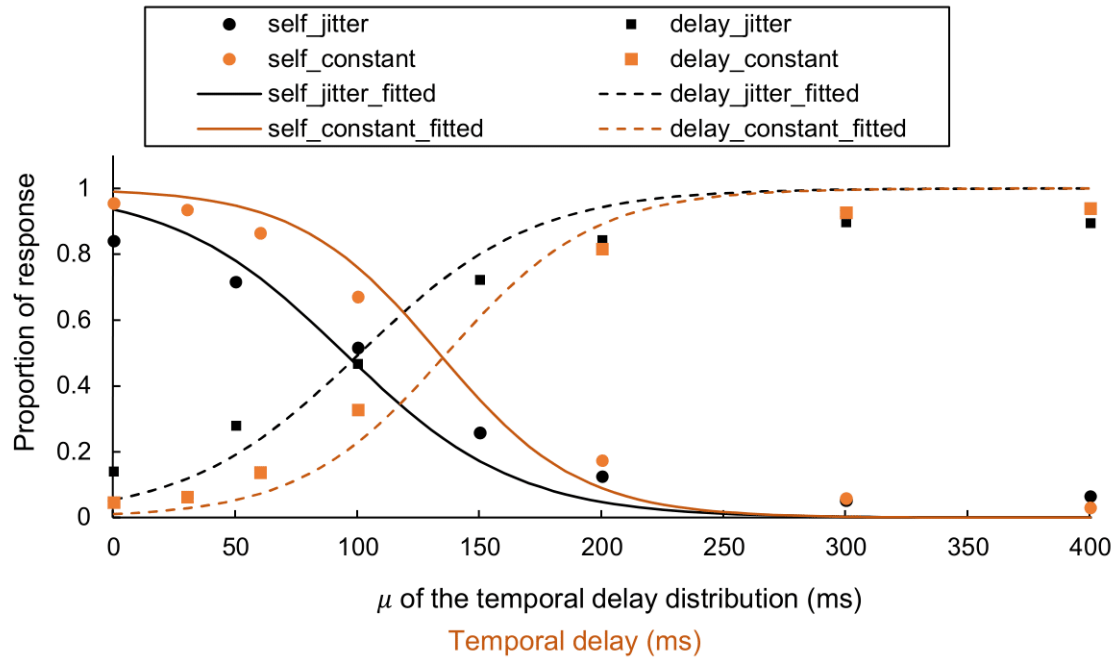


Figure 2.5. For the reported authorship (self versus delay) for both the present experiment ($N = 11$, jitter delays) and the earlier study ($N = 8$, Experiment 1, constant delays), fitted sigmoid function were presented. For visualization purposes, data points from all participants are aggregated to represent a unified measure of reported authorship. Each analysis (i.e., curve fitting) was conducted individually for every participant and condition. Note that ‘self_jitter’ and ‘delay_jitter’ correspond to the average proportions of ‘self’ and ‘delay’ responses shown in Figure 2.4. Reproduced from Kim et al. [103].

Table 2.1. Mean values and standard deviations for each fitted parameter across the responses (“self” and “delay”) for the current study and the previous study (Experiment 1). The parameter b indicates the delay value where the response proportion reaches the 50% threshold, while the parameter a represents the slope at the threshold value (positive value for an increasing slope and negative value for a decreasing slope).

Study	Response	Parameter	Mean	Standard deviation
Current study	self	a	0.0284	0.0104

(jittery delay)		<i>b</i> (ms)	94.34	66.98
	delay	<i>a</i>	-0.0283	0.0114
Previous study (constant delay)		<i>b</i> (ms)	100.74	69.02
	self	<i>a</i>	0.0347	0.0133
		<i>b</i> (ms)	132.97	51.35
	delay	<i>a</i>	-0.0332	0.0141
		<i>b</i> (ms)	136.42	54.36

The authorship rating on a scale of one to six gauges subjective authorship, where one means “I could not possibly manipulate the window” and six means “I fully manipulated the window.” The averages and standard deviations of the ratings were calculated for each participant (Figure 2.6). Mauchly’s test indicated that the assumption of sphericity was violated ($\chi^2[20] = 99.822, p = .000$). Therefore, the degrees of freedom were adjusted using Greenhouse–Geisser estimates of sphericity ($\epsilon = .217$). The main effect of delay was significant ($F [1.303, 13.025] = 65.974, p = .000$, partial $\eta^2 = .868$), indicating that delay affects the subjective authorship of the gaze-contingent window. All parameters for the ANOVA are available in Appendix Table 2.6.

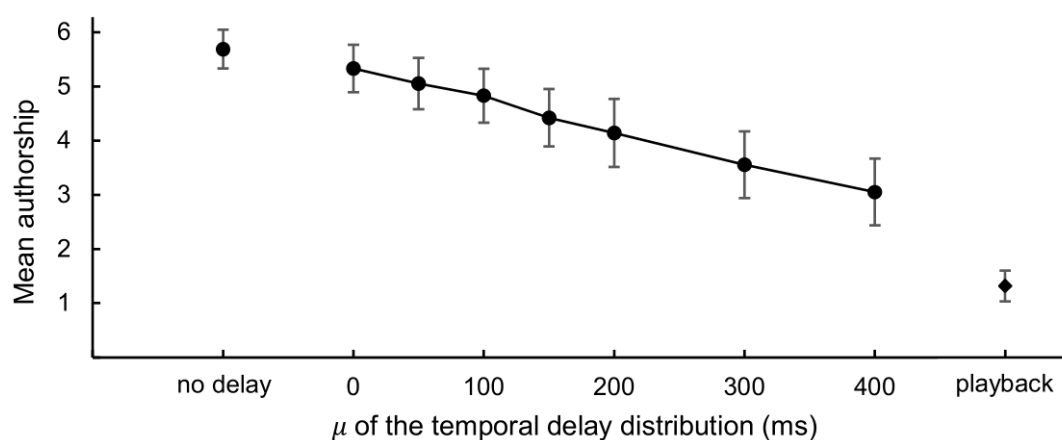


Figure 2.6. Average authorship ratings and its standard deviations across the 11 participants.

Reproduced from Kim et al. [103].

Behaviors

The lowest tasks correct response rate was at 98%, indicating that participants engaged actively in each trial without skipping any (the highest individual error rate was 2.1%; see Appendix 2.2 for details).

Both the mean and standard deviation of the response times were calculated using each participant's average response time under each condition, both with and without the target present (Figure 2.7). Mauchly's test revealed that the assumption of sphericity was violated for both the delay ($\chi^2 [20] = 44.342$, $p = .002$) and the interaction effect ($\chi^2 [20] = 50.915$, $p = .000$). Consequently, degrees of freedom were adjusted using Greenhouse–Geisser corrections (delay: $\varepsilon = .289$, interaction: $\varepsilon = .399$). The analysis showed a significant main effect for delay ($F [1.734, 17.341] = 33.814$, $p < .001$, partial $\eta^2 = .772$) and for the target ($F [1, 10] = 23.471$, $p = .001$, partial $\eta^2 = .701$), while the interaction effect did not reach significance. These results indicate that temporal delay has an impact on response time. All parameters for the ANOVA are available in Appendix Table 2.7. A post hoc pairwise comparison with Bonferroni correction revealed a significant increase in response time by a mean difference of 1,511 ms ($p = .001$) when comparing the target-absent condition ($M = 6,674$ ms) to the target-present condition ($M = 5,163$ ms). This suggests that the current gaze-contingent display necessitates a serial visual search.

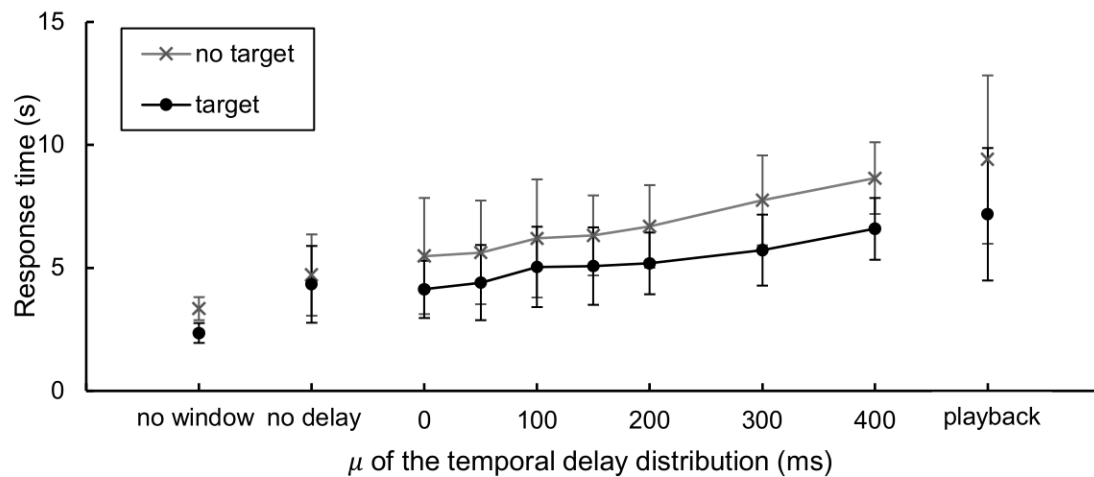


Figure 2.7. Average response time and its standard deviations ($N = 11$). Note that two participants did not perform the 'no window' condition. Reproduced from Kim et al. [103].

Fixation counts per trial were presented for each participant (Figure 2.8), showing considerable variation across participants. Mauchly's test indicated a violation of the sphericity assumption for the delay ($\chi^2[20] = 53.481, p = .000$); thus, degrees of freedom were adjusted using Greenhouse–Geisser estimates ($\epsilon = .292$). Main effect was not significant for Delay ($F [1.751, 17.512] = .774, p = .460$, partial $\eta^2 = .072$). All parameters for the ANOVA are available in Appendix Table 2.8.

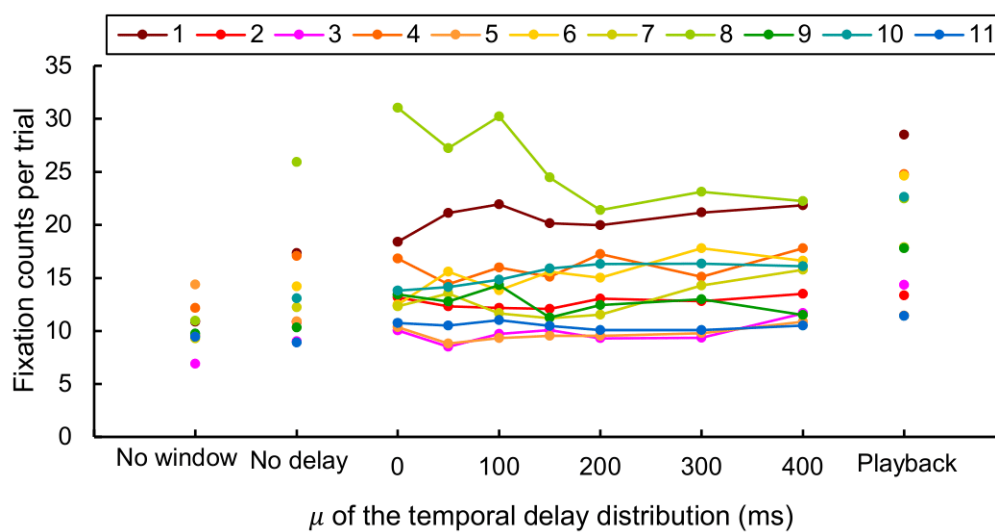


Figure 2.8. Fixation counts per trial with 11 participants. Note that two participants did not perform the 'no window' condition. Reproduced from Kim et al. [103].

The average fixation count per second, along with standard deviations, was calculated for both the current experiment involving a jitter delay ($N = 11$) and the previous study with a constant delay ($N = 8$, Experiment 1) (see Figure 2.9). The type of experiment did not result in a significant main effect ($F [1, 17] = .961$, $p = .341$, partial $\eta^2 = .053$). However, delay exhibited a significant main effect across both experiments ($F [1.473, 25.034] = 90.652$, $p < .001$, partial $\eta^2 = .842$). Mauchly's test indicated a violation of the sphericity assumption ($\chi^2[9] = 52.956$, $p < .001$), leading to an adjustment of the degrees of freedom using Greenhouse-Geisser estimates ($\epsilon = .368$). Detailed parameters for the ANOVA are provided in Appendix Table 2.9.

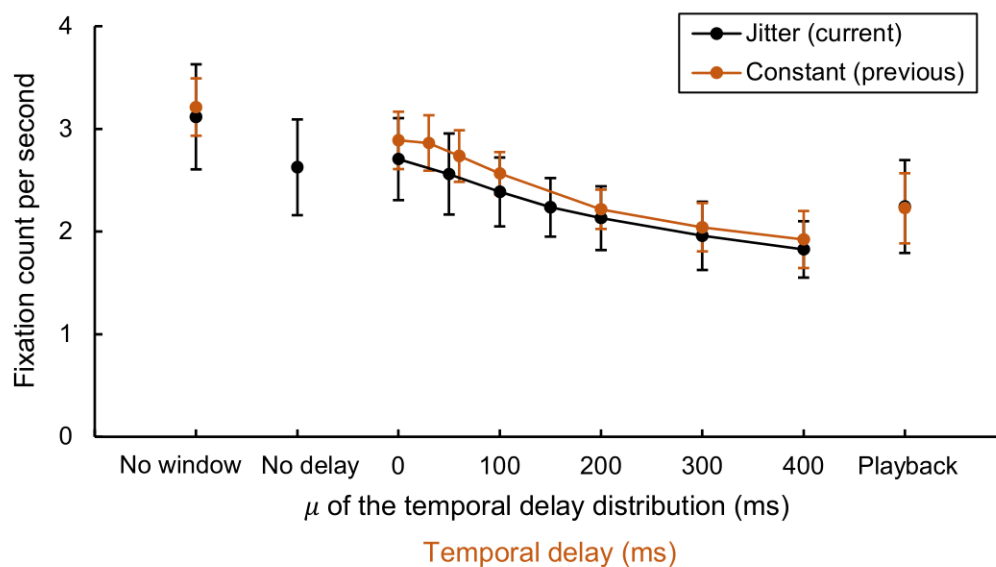


Figure 2.9. Average fixation count per second was calculated for both the current study, which involved a jitter delay ($N = 11$), and the previous study, which utilized a constant delay ($N = 8$, Experiment 1). Mean values among participants are displayed with standard deviation represented as error bars. It is important to note that the 'no delay' condition in the current experiment is virtually identical to the '0 ms' condition in the previous experiment, as no intentional delay was introduced. Reproduced from Kim et al. [103].

The mean relative distribution of fixation durations was calculated for the 11 participants (Figure 2.10). In the no-window condition, the distribution mode appeared at 150 ms. Given that the no-delay and playback conditions had similar modes at 200 ms, this seems attributable to the gaze-contingent window. With the delay, the mode shifted to 225 ms. Notably, when the μ was 50 ms, I observed dual modes: one remained around 125–225 ms, and another mode appeared at the position of the sum of μ and 200 ms (e.g., when the μ was 300 ms, the latter mode was at 500 ms).

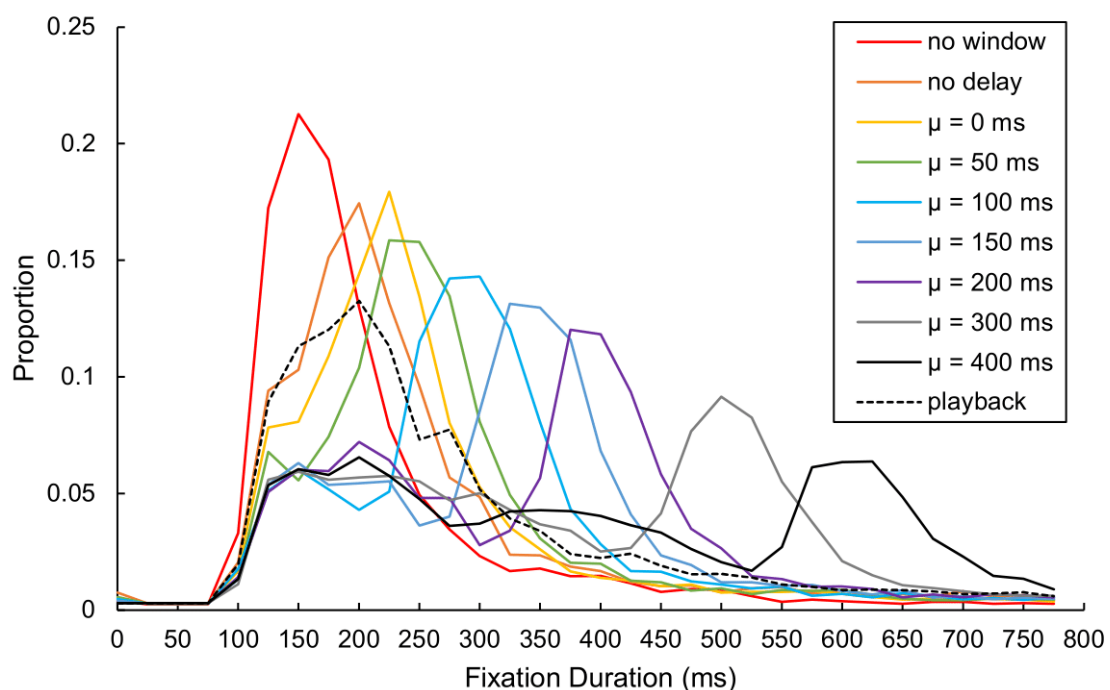


Figure 2.10. Relative distribution of fixation duration. Each participant's fixation duration within every 25 ms bin, ranging from 0 to 800 ms, was calculated. The graph presents the average values calculated across participants, with each bin's representative value placed at its minimum (e.g., the value of the 200–225 ms bin is shown at 200 ms on the horizontal axis) (N = 11). Note that two participants did not perform the 'no window' condition.

Reproduced from Kim et al. [103].

The mean relative distribution of saccade amplitudes was again calculated for the 11 participants (Figure 2.11). Across all conditions, two distinct modes were observed: one at 1° and another at 7°. Given that the shortest distance between two characters was 8°, this likely explains the peak at 7°. The relative frequency below 1° seemed to increase with the gaze-contingent window when comparing no delay or playback to the no-window condition, and it further increased with the μ .

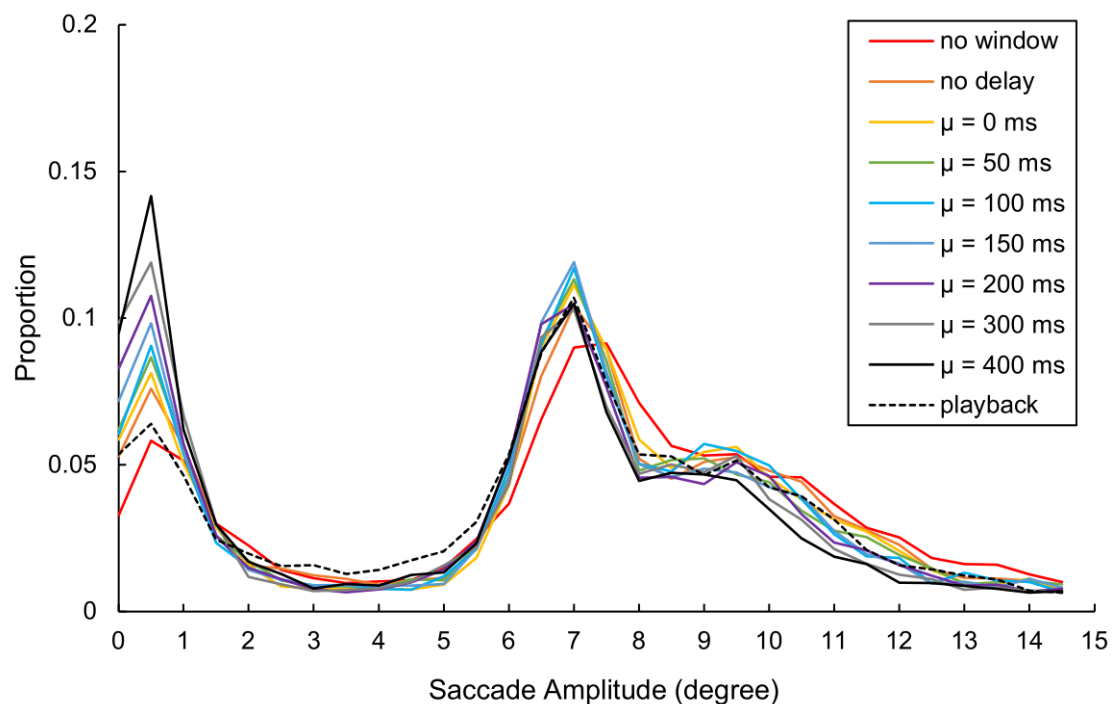


Figure 2.11. Relative distribution of saccade amplitude. Each participant's saccade amplitude within every 0.5° bin, ranging from 0 to 15°, was calculated. The graph presents the average values calculated across participants, with each bin's representative value placed at its minimum (e.g., the value of the 10-10.5° bin is shown at 10° on the horizontal axis) (N = 11). Note that two participants did not perform the 'no window' condition. Reproduced from Kim et al. [103].

Discussion

This study aimed to investigate the impact of jittery delay on the sense of agency and behaviors, as well as to identify the statistical attributes that most effectively explain the

relationship between the sense of agency and eye movements under jittery delay conditions.

Influences on the sense of agency

Categorical authorship reports showed that self-responses decreased as the delay increased. The separation between ‘delay’ response frequency and response time in playback conditions, along with the predominance of other responses in the conditions, suggests that ‘delay’ responses are more closely related to the subjective sense of agency than task difficulty. Similarly, the consistent authorship rating between one and two for playback conditions suggests that these ratings likely reflect participants' perception of the authoring agent rather than task difficulty. The predominantly reported delay response (i.e., partial sense of agency) led to a more significant decrease in the proportion of self-responses than the decrease in the authorship rating, where a rating of one implied a total absence of the sense of agency. The decrease in authorship ratings with temporal delay aligned well with prior research that used hand movements as the primary action [18,21–24]. Specifically, the current categorical authorship reports align with those of Farrer et al. [19] in two key aspects: 1) self-responses decreased while delay responses increased with temporal discrepancy, and 2) other responses were seldom reported, even at longer temporal discrepancies.

However, Farrer et al. [20]'s study found participants misattributing authorship and choosing other responses at longer temporal discrepancies. These misattributions can be explained by the degrees of freedom in the action-feedback experimental paradigm. Both the current study and Farrer et al. [19]'s study allowed participants to incorporate movement direction into their authorship judgments (gaze direction and joystick manipulation, respectively). Farrer et al. [20]'s study provided only a single degree of freedom (a single-button press), making it difficult for participants to claim authorship at extended temporal discrepancies. Such variations in sensory data across studies might further explain the differences in the attenuation degree of subjective reports. Farrer et al. [20]'s study

documented that an increase in the partial control response proportion started at longer temporal discrepancies than in the current study. Wen [34] suggested that abundant sensory data would narrow the time window for a sense of agency. These differences do not significantly undermine the current claim that this study generally replicates prior research, even using eye movement as the action modality.

Influences on behaviors

The fixation counts per trial remained consistent regardless of the μ value, while the fixation counts per second steadily decreased. This implies that the number of fixations necessary to complete the visual search did not vary across different delay conditions, even though the delays impacted their visual search time efficiency. This is further supported by the increase in response times. The decline in fixation counts per second can be clarified by analyzing the fixation duration distributions. These distributions reveal two distinct patterns: 1) a shifting mode (referred to as the 'higher mode') and 2) a consistent mode between 125 and 225 ms (referred to as the 'lower mode'). The location of the higher mode closely aligns with the delay, shifting in accordance with the μ value. Subtracting the μ value from the position of each higher mode across all delay conditions shows that all mode positions converge at 200 ms, except when μ is 0 ms, where it converges at 225 ms. This correlation likely reflects participant behavior of waiting for the gaze-contingent window to update at their fixation point before reading the character, termed 'wait-and-read' for the higher mode, and identified as typical reading behavior for the lower mode.

The distributions of saccade amplitudes indicated an effect of delay, as shown by an increase in the relative frequency of the 0-1° saccades. While I remain cautious about the current camera-based eye-tracking capability to accurately record microsaccades, I can infer the underlying causes from previous research on microsaccades. Prior studies have suggested that microsaccades can serve as a proxy for covert attention allocation [90,91]. I propose that

the separation between eye movement and attention allocation explains the results observed with the current delayed gaze-contingent display. When the window follows eye movements with a delay, it creates a gap between the participant's gaze point on the monitor and the window. This gap may require participants to direct their attention to the window as part of their temporal mismatch perceptions while keeping their gaze on the desired character to perform a visual search. With increasing delay, widening the gap between the window and their eye movement, the frequency of covertly attending to the window increases.

Although the action-effect delay did not significantly hinder the participants' visual search, their attentional resources were indeed utilized. Despite the low error rates and the negligible impact of the μ value on fixation count per trial indicating a minimal effect of the delay on search behaviors, the variation in saccade amplitudes in response to μ values suggests an adaptive attention distribution that requires further exploration. Reduced visual search efficiency implies higher attentional demands on participants' working memory. The use of two fixation strategies indicates that participants chose to 'wait and read' to reduce these demands. These eye movement results suggest a possible proportional relationship between the amount of attention consumed and the length of the delay. However, this relationship cannot be confirmed as the delay did not impact task performance. Although attention plays a critical role in comparing actions and their effects [94] and is a finite resource necessary for the experience of agency [106], the current study, which demonstrated sustained search performance and a steady decline in subjective authorship, suggests that visual search and the sense of agency adequately shared the available attentional load. These findings underscore the importance of further research on task difficulty.

Statistical attributes of jitter on the sense of agency

In comparing constant delay conditions from Experiment 1, no significant differences were observed in the subjective report parameters or the fixation counts per second.

Additionally, the relative distributions of fixation durations and saccade amplitudes showed no significant impacts. These results suggest that the minimum value of the distribution primarily influences the sense of agency and eye movements. When generalizing the current findings, qualitative differences in fitted values should be taken into account. The plots of fitted values per participant (see Appendix Figures 2.2 and 2.3) suggest that jittery delay might lead to less uncertainty and shorter delays to impair the sense of agency. Further research is needed to confirm the conclusions. These conclusions are also limited by the current choice of distribution, with σ set at 50 ms.

Direct comparisons with prior studies are limited due to methodological differences, such as continuous action–effect interactions and the use of jittery distributions across delay ranges. Normoyle et al. [107] provide a somewhat relevant comparison by adjusting the θ value of a gamma distribution to achieve variances of 0, 50, 100, and 150 ms, while maintaining a constant mean of 200 ms. In their study, participants only perceived delays at the highest variance of 150 ms. The current study used a truncated normal distribution with a variance of 628.29 ms across all delay conditions, identifying a subjective report threshold of 50% at a delay of 94 ms, which is the distribution's minimum value, with a mean delay of 130.14 ms. Although comparing directly with gamma distributions poses challenges, the current findings reveal a greater variance in threshold delay (average = 130 ms, variance = 628 ms, lowest value = 94 ms, highest value = 194 ms) compared to earlier studies (average = 200 ms, variance = 150 ms). Previous studies probably experienced rare longer delays (e.g., surpassing 400 ms), which may account for the observed temporal delay with a smaller variance. This difference suggests that perceptions of temporal mismatch may be more influenced by delays near the distribution's minimum rather than its maximum value, underscoring the need for future research to encompass a wider range of delays.

These experiments suggest that the threshold for perceiving delays may depend on

both the length of the delay and its frequency of occurrence. Sporadic extended delays are critical when considering data packet transmission over the internet; an extended delay in one packet often predicts similar delays in subsequent packets [108]. Additionally, packet loss often occurs in clusters, impacting consecutive packets [109]. In continuous action-effect human-computer interaction (HCI), even brief instances of such sporadic, consecutive extended delays can severely disrupt the sense of agency and potentially modify user behavior. Consequently, future research should include experiments with a broader range of delays and explore the effects of different frequencies of these delays.

Gaze-contingent display

Low error rates and significant response time differences in target presence conditions confirm that participants navigated the visual search in a serial manner without any skipping, encountering the intended disruption caused by jittery delays. The effects of delays on subjective reports and behaviors validate the current design approach and underscore the efficacy of jittery delays. Additionally, I observed no decrease in overall saccade amplitude, consistently recording 6-8° saccades across all conditions. The 3×3 virtual grid used in the current experiment likely prevented reductions in saccade amplitude caused by the gaze-contingent window, distinguishing the current methodology from previous research where saccade amplitudes decreased with peripheral blurring gaze-contingent windows [76,78–80,82]. Thus, the eye-gaze interface implemented was successful in reducing the impact of the window, enabling a concentrated investigation of the effects of delay.

Conclusion

Chapter 3 extended our understanding by exploring the effects of jittery delays on subjective authorship and eye movements. The current results again confirmed a relationship between the delay and the refinement of strategies, supporting the compactor model as a underlying mechanism in the sense of agency using eye movements. As the first systematic

exploration of jittery delays on the sense of agency, Chapter 3 expands our knowledge of eye-gaze HCI in environments characterized by delay variability. Chapter 3 revealed that variability of up to 100 ms in delay does not affect the sense of agency and task strategies, which informs guidelines for designing interfaces resilient to network variability, ensuring consistent and reliable user interaction. Moreover, this experiment again highlighted the potential of eye-gaze HCI to detect alterations in gaze behavior as indicators of users' delay perception.

Chapter 4

Chapter 2 suggests that the sense of agency in eye-gaze HCI emerges from an internal comparison of one's actions and their effects despite a lack of direct support and only theoretical plausibility. In this experiment, I aim to identify brain activations associated with the sense of agency in eye-gaze HCI using functional magnetic resonance imaging (fMRI).

In the previous research where such brain activity had been investigated with hand movements, the right inferior parietal lobule (IPL), especially the right angular gyrus (rAg), was reported in both studies to be negatively correlated with the magnitude of the discrepancy between action and the following effect [37,110]. The parietal cortex is posited to oversee intentions and motor plans, actively identifying when the outcomes of actions align with their intended goals, as highlighted in the work by Sirigu et al. [111]. Farrer et al. [37] suggest that this area plays a critical role in motor control, enabling the conscious recognition of one's actions, particularly those tied to intentions and the sense of agency. Based on the previous observation, I hypothesize that the discrepancy between the intentional motor signal of eye movements and the outcome is perceived by the activation in this area, thus contributing to the sense of agency, particularly in the right inferior parietal lobule and right angular gyrus, resulting in a resemblance between the current experiment and the previous studies [37,110]. The current experiment sought to investigate the following: 1) delay-modulated brain activation and 2) neural activation related to the perception of action author (self vs. other).

The previous findings from Chapters 2 and 3 confirmed that in the eye-gaze HCI's delay from 0 to 300 ms, the delay has a near-linear proportional relationship with the proportion of "delay" response to the categorical questionnaire. Thus, by implementing the delay condition of 0, 100, and 300 ms in the current experiment, here I consider the independent variable of the delay as a proxy that represents the participants' delay perception.

Additionally, I added a 1000 ms delayed condition to the current experiment to ensure that participants perceive the delay in the newly adopted eye-gaze HCI. Finally, an uncontrollable condition was added for participants to judge the computer as the event author consistently.

As the influence of delay on eye movements was not the objective of the current experiment, the current experiment constrains the participants' eye movements during the visual search. Specifically, participants were allowed to execute a saccade only once during a visual search trial. This was to exclude any impact of the delay on the variance of their eye movement behavior and further their brain activity related to the variance of the search behavior. Following the eye movement restriction, a visual search trial was restricted to 2 s.

Since the saccade was restricted to once, in the "playback" condition, the gaze-contingent window may move to the exact location as a participant's real-time saccade target location. Furthermore, the playing back window may move as it follows the participant's real-time gaze with a delay. Thus, I implemented the uncontrollable condition with a new method. In the uncontrollable condition, the window always moved to a location other than the participant's gaze. As the participants' input action was limited to only once, any changes in the stimuli can be perceived by the participants as follows: "I controlled the stimuli, but the update was delayed." Hence, in the uncontrollable condition, the window's movement had a random delay (300 or 1000 ms) from the participants' real-time gaze, so the participant could not predict the delay. Furthermore, the participants were instructed that the computer occasionally controls the interface instead of them, giving them a contextual cue on this incongruency condition. This strengthened the perceived unpredictability in this condition, thus ensuring participants perceive that the "computer controls the stimuli but not myself."

The block design approach was adopted instead of an event-related design, as an event-related design is considered to be impractical for this experiment. It would be necessary to jitter the interstimulus interval (ISI) to implement this experiment with a randomized

event-related design considering the effects of anticipation and the hemodynamic response function (HRF) [112]. However, this would also introduce jittering to the condition of temporal incongruency in the gaze-contingent window's update. If participants are required to report the sense of agency with a complex mix of jittering in both ISI and condition delay, finely tuned to the time flow throughout a session, it could introduce confounding factors that disrupt their belief system. This experiment fundamentally relies on the participants' trust in the computer, that the window will often follow their eye-gaze's spatial location with various delays and barely does not (in the incongruency condition). The participant's perception of the sense of agency depends on their belief system constructed by the experiment instructions and practice sessions (a preliminary experiment was conducted with the same experimental setup before the fMRI experiment; refer to the Procedure and Task section). The belief is a crucial component in the conceptual judgment of agency from the two-component model proposed by Synofzik et al. [26]. Based on the participants' belief, the temporal and spatial discrepancy in the incongruency condition are heterogeneous elements. While the current experiment places temporal and spatial incongruencies on the same scale (i.e., sense of agency), this is made possible by the block design and is not directly related to the reported higher signal-to-noise ratio than the event-related paradigm [113]. An event-related design could confuse participants, especially those with weaker belief systems. Even if participants have a clear awareness of the computer's agency in the incongruency condition, their ability to interpret spatial incongruency as an intentional action by the computer rather than a bug depends on the strength of their belief. Eventually, this belief could be shaken in an event-related design where delays are jittered, and the window's position has a spatial discrepancy, significantly if the timing of both temporal and spatial discrepancy is influenced by preceding item conditions (i.e., condition type or ISI), leading to potential confusion.

Materials and Methods

Participants

Initially, 28 participants were recruited for the study. However, five participants were excluded for the following reasons: 1. Eye tracking during fMRI could not be maintained. 2. Declined based on the MRI safety guidelines. Data from 23 participants were analyzed for the current study, who were all students (19 undergraduates, 4 pursuing master's degrees) at the Tokyo Institute of Technology. The participants had an average age of 20.6 years, with a standard deviation of 2.4 and an age range of 19–25 years. Among the participants, 4 were women, including 20 Japanese, 3 Korean, and 1 Chinese. This study was introduced to undergraduate students during their classes, while master's program students were notified through their advisors via email. Participants received 10,000 yen upon completion of the experiment. All participants had either normal or corrected vision and gave informed consent prior to taking part the experiment. The experiments adhered to the guidelines of the Declaration of Helsinki. Institutional consent for publication was obtained from all participants. The Tokyo Institute of Technology Ethics Review Committee for Epidemiological Research approved the study. Data collection occurred from March 2 to 15, 2023, followed by statistical analyses and evaluations to address the research questions.

Procedure and Task

Before beginning the experiments, the participants were instructed that a computer could delay or manipulate the display throughout a block. Each participant completed four sessions, totaling 80 blocks. Every session comprised 20 task blocks, including five trials (Figure 3.1). During each 2-second trial, participants engaged in a constrained visual search task on a bi-resolution gaze-contingent display, which blurs the periphery (see the detailed description in the Stimuli section). Participants were instructed to focus on the center character at the start of each trial and then make a single saccade to one of the four surrounding characters (the surrounding four were the same). They were advised to select

each of the four locations as evenly as possible across trials. Participants were to maintain their gaze on the initially fixated character until the end of the trial. They were required to press a button if the center character matched any of the following characters and refrain from pressing any button if the characters differed. The surrounding characters could either match or differ from the center character. I instructed the participants that the computer might occasionally control the gaze-contingent window when they could not. In this condition, after the participant's saccade, they were asked to halt their gaze even when the windows were presented in another location, and they could only see the blurred character in their central vision. Also, the participants were allowed to judge whether the unblurred character at their periphery was the same as the center character. After completing five visual searches under the same condition (16 seconds in total), participants reported whether the display update was self-initiated (hereafter referred to as self-response), delayed (delay response), or appeared to be manipulated by another agent (other response, these were the categorical questionnaire). Subsequently, a rest block, including the time for the authorship report, totaled 14 seconds.

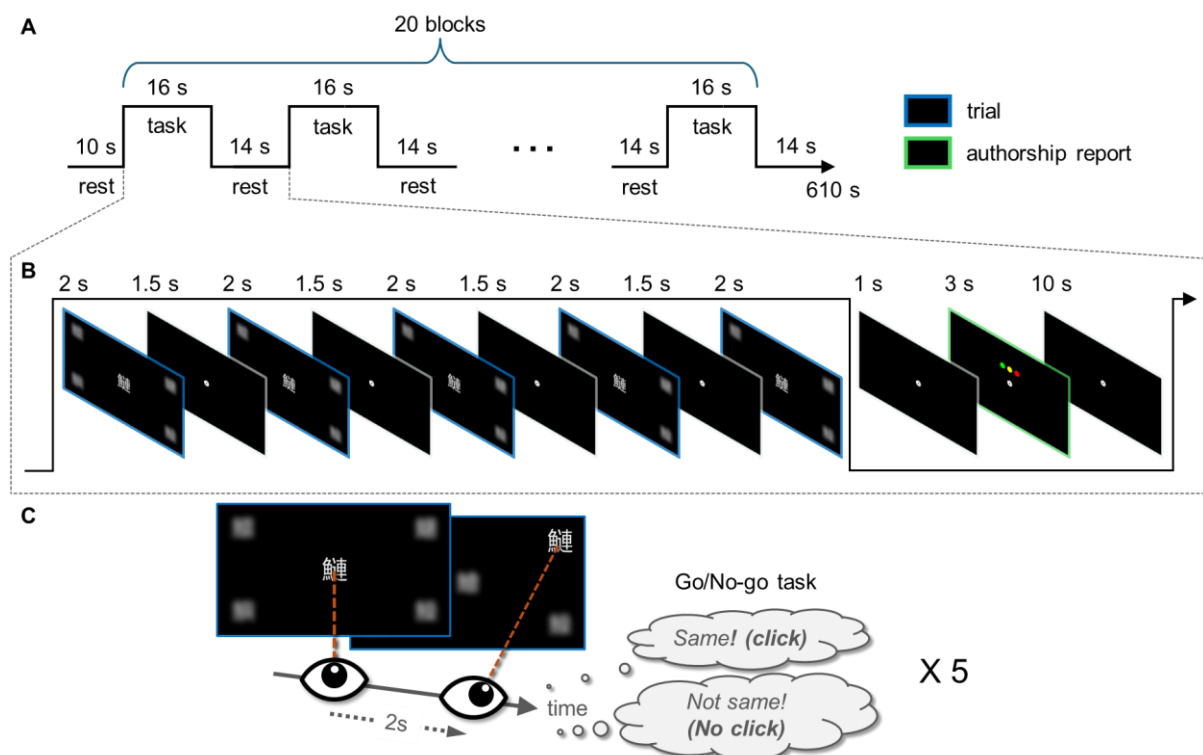


Figure 3.1. Block design representing a session: (A) Each session comprises 20 task blocks. (B) Each block contains five trials. In every 2-second trial, the task involves presenting a central character alongside surrounding characters, identical to or different from the central character. After the block, subjective authorship was reported and rested, comprising 14 seconds. Additionally, within each block, the condition is randomly selected and uniform throughout the block. (C) In each trial, the participant had to initiate a saccade from the center to one of four surrounding locations. Then click the response button only when the two characters are identical. Subsequently, they had to maintain the gaze location until the trial was terminated (2 s).

Before the main experiment and without MRI, a preliminary session with the same experimental setup was conducted in a darkened room, but the sessions were executed in a different sequence. This preliminary phase served not only for practice but also to validate participants' performance by comparing it across sessions, ensuring the reliability of the experimental data. I validated all the behavioral data taken with the main experiment.

Conditions

Each session comprised four repetitions of five conditions: four delayed conditions (0, 100, 300, and 1000 ms) and a temporal/spatial incongruency condition (herein, I refer to this condition as the incongruency condition). In delayed trials, the window had a temporal delay between the participant's eye movement and the display update. The incongruency condition was designed to prevent participants from anticipating the feedback, creating an experience where an external agent controlled the display. In this condition, 300 and 1000-ms delays were pseudo-randomly set within a block so that participants could not predict the delay. Then, the window mirrored the gaze position in the X or Y axis. Both axes pass along the center position of the monitor. The presenting sequence of the five conditions was predefined with T1I1 to counter carry-over effects [114]. In every task block, the center and surrounding

characters were the same in two trials and different in three. Presenting sequence was predefined with T1I1. T1I1 sequencing was also used in the presenting sequence of the 300 or 1000-ms temporal incongruency and x-axis or y-axis spatial incongruency in five consecutive visual searches in a block of the incongruency condition.

Stimuli

On a visual search array (Figure 3.2), each Chinese character had a size of 2°. The surrounding four characters stay apart 15.5 degrees horizontally and 7 degrees vertically. Compared to Chapters 2 and 3, the display only differs in size; otherwise, it shares the same rendering method.

With 239 Chinese characters (see Appendix 1.1), 400 images were created (100 per session). The center and surrounding characters shared the stroke counts and the meaning of fish by having the same radical, thus having visual similarity. This prevented meaning- or category-based attentional bias and controlled processing time per character within a trial. The presenting sequence of the characters was defined with T1I1.

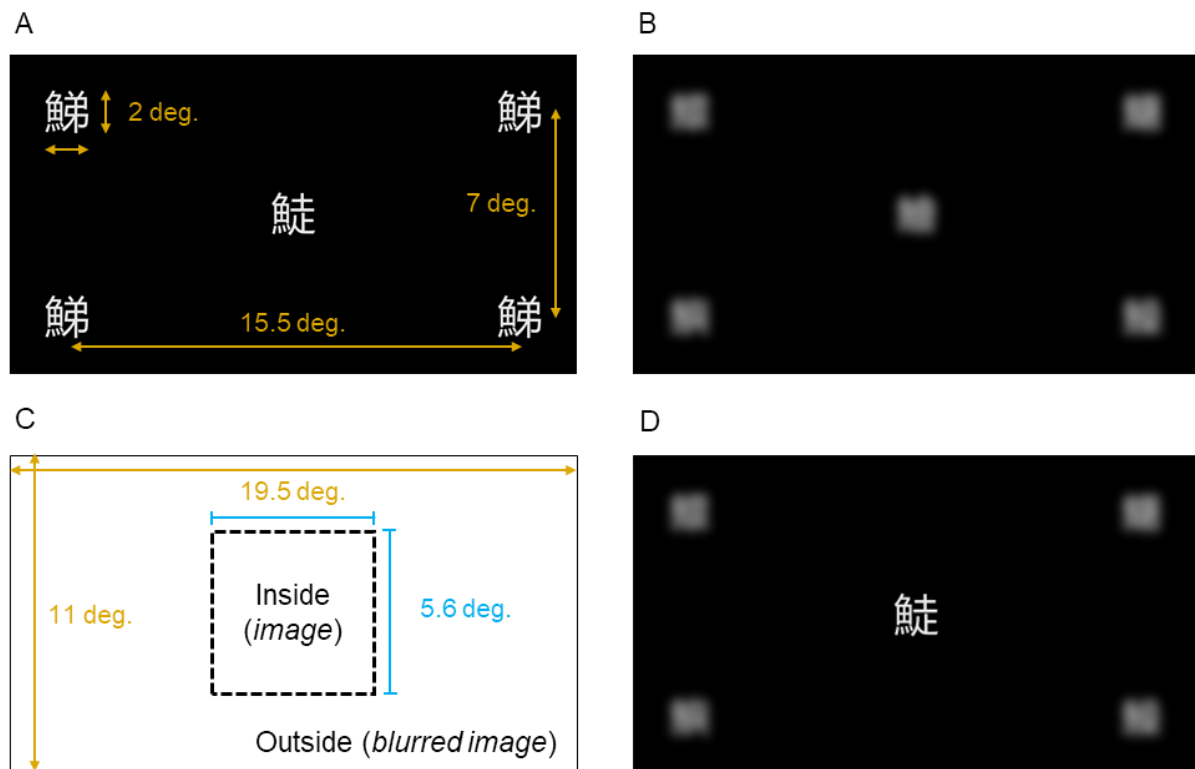


Figure 3.2. A, top-left: A stimulus of five Chinese characters arranged in the center, and a two-by-two virtual matrix was generated randomly. All five characters had the exact stroke count and included a fish radical. A total of 540 images were created, with the target character either included or excluded. B, top-right: One of the 540 generated images was randomly selected and blurred. C, bottom-left: Within the gaze-contingent display (rectangle with solid borders), a virtual window following the gaze (square with dashed borders) was centered on the participant's fixation location. The image was rendered inside this window, with the surrounding area blurred. D, bottom-right: An example of the stimulus when the participant focused their gaze on the central Chinese character.

Apparatus and Magnetic Resonance Imaging Data Acquisition

A 3-T SIEMENS PRISMA scanner with a SIEMENS 32-channel head coil was used for both structural and functional imaging. The use of a foam pillow and padding minimized head movements. For each functional run, a T2*-weighted gradient-echo echo-planar imaging (EPI) sequence sensitive to blood-oxygenation-level-dependent (BOLD) contrast was used to acquire 36 slices covering the whole brain, in descending order, per time repetition (TR). The following parameters were used: 0.75 mm slice gap, $3 \times 3 \times 3$ mm resolution over a $192 \times 192 \times 134.25$ mm FOV, time echo (TE)/TR = 30/2000 ms; flip angle = 77 degrees. For the T1-weighted structural image, magnetization-prepared rapid gradient echo (MPRAGE) was acquired with the following parameters: $1 \times 1 \times 1$ mm resolution over a $256 \times 256 \times 192$ mm FOV; inversion time (TI)/ TE/TR = 900/2.98/2300 ms; flip angle = 9 degrees.

BOLDscreen 32 LCD for fMRI (Cambridge Research Systems, 32-inch, 120 Hz, 1920×1080) was used to present the current stimuli. Pyka (CURRENT DESIGNS, five-button fiber optic response pad) was used to obtain participants' responses. Left eye movements were recorded monocularly in a proprietary PC and simultaneously sent to

another PC for updating the stimuli with an EyeLink 1000 Plus eye tracking system (SR Research Ltd.), with a spatial resolution of $<0.02^\circ$ and a sampling rate of 1,000 Hz.

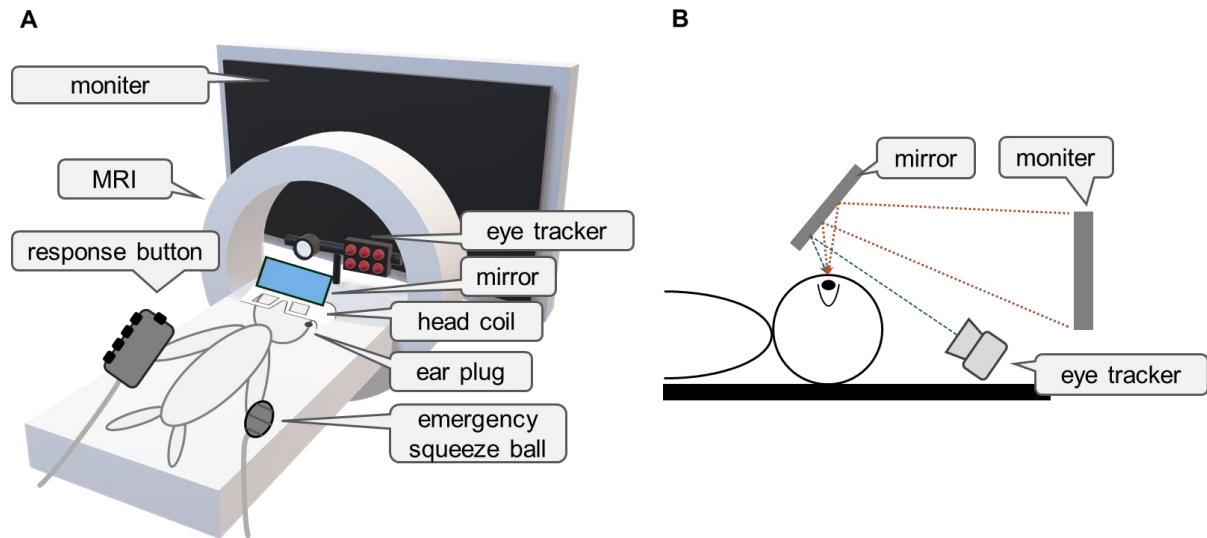


Figure 3.3. Experimental setup. A) Participants have a response button and emergency pneumatic squeeze ball. B) Eye tracking was established with an eye tracker directed to the mirror.

Preprocessing of Functional Magnetic Resonance Imaging Data

Image and statistical analyses were performed using Statistical Parametric Mapping (SPM12, version 7771, www.fil.ion.ucl.ac.uk/spm/software/spm12). Before each functional run, five images were acquired and discarded to allow longitudinal magnetization to approach equilibrium. The image volumes were first temporally realigned to correct for the order of slice acquisition and then spatially realigned to the initial image to compensate for artifacts related to motions. Slice timing correction within each volume was performed with the 18th slice as a reference. Head movements in a session were corrected with a six-parameter (rigid-body) spatial transformation. The T1 anatomical image was coregistered with the averaged functional image generated in the previous step. The functional images underwent normalization with the template (T1 template of the Montreal Neurological Institute space)

with the voxel dimensions resampled to $2 \times 2 \times 2 \text{ mm}^3$ [115]. Finally, spatial smoothing of the images was carried out using a Gaussian kernel with an 8 mm full-width at half maximum.

Individual Analysis

I employed a consistent set of analysis techniques for the present study across the three analyses. These common methods are outlined here, while specific differences for each analysis are described in the following sections.

A design matrix was created for each subject encompassing the four sessions of the visual search task. The fMRI data was analyzed using a general linear model (GLM) for each subject [116]. Neural activity during the task was modeled with box-car functions convolved with the canonical hemodynamic-response function. Throughout the analyses, the questionnaire regressor was created as a regressor of no interest. The various task-related regressors used in each analysis are detailed in the following sections. For all analyses, the time series of each voxel underwent high-pass filtering at 1/128 Hz. In assuming a first-order autoregressive model, serial correlation was estimated using the restricted maximum likelihood procedure, which was then applied to whiten the data and design matrix [117]. To minimize motion-related noises, six rigid body head motion parameters were incorporated as regressors of no interest [118].

Group Analysis

Here, consistent methods throughout the three analyses were described. Throughout the three group-level analyses, for the height threshold, family-wise error (FWE) correction was applied to control for multiple comparisons, with corrected p-values below 0.05. Clusters smaller than eight voxels (64 mm^3) were excluded from the results. Brain regions were identified and labeled based on the atlas Neuromorphometric labels [119] under academic subscription. The brain regions were also identified using Brodmann areas and the Talairach atlas using the Talairach client (<https://www.talairach.org/index.html>) [120,121]. MNI

coordinates were converted to Talairach coordinates using the transformation tool Lacadie et al. provided[76]. The result tables present the coordinates in MNI space.

Delay-Modulated Analysis

As I hypothesized that the activity of the rAg increases with the delay, I adopted the magnitude of delay in millisecond units as a parametric variate. The main contrast of interest was the parametric effect of delay in milliseconds. Each session included two task-related regressors. The two task-related regressors were 1) delayed condition and 2) incongruency condition. In the regressor of delayed condition, the magnitude of delay in millisecond units was set as a parametric variate. For group-level analysis, individual contrast images for the parametric effect of delay were subjected to a one-sample t-test to determine group effects.

Delay-Related Analysis

As delay-modulated beta-weights might lead to no voxels surviving, I broadened the search area by conducting a repeated ANOVA analysis. The GLM included five task-related regressors. The five task-related regressors were for each condition (0, 100, 300, 1000 ms, and incongruency condition). A one-way ANOVA was conducted on the four individual-level contrasts using repeated measures. Thus, each delay condition is compared against the incongruency condition for group-level analysis. This approach was taken to narrow the scope of the investigation solely to the participants' subjective feelings of authorship. By comparing each delay condition to the incongruency condition, I aimed to highlight only the differences in the perceived authorship from self to other (i.e., computer), thereby eliminating irrelevant activity. The ANOVA was to identify differences between condition levels without specifying the direction of the effect. The measurements in the four delay condition levels were treated as interdependent and had equal variance.

As the ANOVA analysis revealed a voxel cluster located at the right supramarginal gyrus (rSMG, refer to the Results section), I further investigated the degree of activation (i.e.,

mean beta weights) at each condition. Using the “Leave-One-Subject-Out” (LOSO) method [122], a region of interest (ROI) was created manually by selecting a voxel cluster (FWE-corrected p-value under 0.05) revealed as significant in the ANOVA analysis with the data of 22 subjects, excluding one subject, located at the supramarginal gyrus (MNI coordinates approximately 60, -42, 28). ROI was created using the MarsBar toolbox [123]. For the excluded subject, the beta values within this ROI were averaged across voxels and time under each of the four delayed conditions. This method calculated mean beta weights for each participant and each delayed condition. This method allowed for an unbiased assessment of the activation patterns by systematically excluding each subject one at a time from the analysis, thereby preventing effect size inflation.

Incongruent vs. 0 ms Analysis

This contrast aimed to identify brain activation regions specifically associated with the participants’ perceived inability to control the gaze-contingent display. The GLM included five task-related regressors, the same as the delay-related analysis. The main contrast of interest was comparing the incongruent and the 0 ms conditions. For group-level analysis, individual contrast images of the “incongruent - 0 ms” comparison were subjected to a one-sample t-test.

Behavioral Data Analysis

I calculated the mean correct rate for each condition, assuming it would be above the chance level (50%) in the delayed conditions and near the chance level in the incongruency condition. Since the window did not follow the participants’ gaze in the incongruency condition, I assumed they could not conduct the visual search. Hence, the correct rate is near the chance level. The mean correct rate and its standard deviation were calculated among the participants. Additionally, I calculated the mean reaction time for each condition, assuming that it would be under 2 s, which is the time when the gaze-contingent stimuli were presented.

The mean reaction time and its standard deviation were calculated across the participants.

For categorical response, I utilized two generalized linear mixed models (GLMM) with SPSS 24.0, excluding incongruency condition trials, to compare “self” versus “delay” responses and “other” versus “delay” responses. These models used a multinomial probability distribution and a generalized logit link function. Participants were included as random effects with variance components as the covariance structure. Fixed effects included the temporal delay. Due to data loss in the subjective report, three responses were excluded, resulting in 1,469 responses being analyzed. The Satterthwaite Formula was used to approximate the degrees of freedom, accounting for the unbalanced data from partial data loss.

Additionally, I investigated the eye movements, focusing on the distribution of fixation duration and saccade amplitude. As I instructed the participants to execute a saccade only once and hold their gaze throughout the experiment, I assumed their eye movements were unaffected by delay conditions. Saccades and fixations were identified from raw gaze data using SR Research Data Viewer software, with predefined minimum thresholds for eye movements (0.1° , $30.0^\circ/\text{s}$, and $8000.0^\circ/\text{s}^2$) set to detect saccades. Fixations outside the stimuli and those occurring immediately before or after a blink were excluded from the analysis. Neighboring fixations within 1 degree and those shorter than 120 ms were merged into a single fixation. The mean fixation duration frequencies were calculated from the representative frequencies of the eight participants. The frequency of fixation durations within each 50 ms bin in the 0–1050 ms range was calculated for each participant and condition. Similarly, the mean relative saccade amplitude frequencies were computed using the representative frequencies of the eight participants. For each participant, the frequency of saccade amplitudes within each 0.5° bin in the 0– 11.5° range was calculated for each condition.

Results

Behavioral Results

Figure 3.4A illustrates the correct rates and reaction times for each condition. The categorical authorship report includes three response levels: self, delay, and other. For each condition, the mean and standard deviation of the proportion for each response type were computed and are depicted in Figure 3.4B. The generalized linear mixed model (GLMM) results indicated a significant main effect of delay when comparing “self” and “delay” responses ($b = -0.006$, $SE = 0.000$, $t = -19.982$, $p < 0.001$, Odds Ratio [OR] = 0.994), showing that as μ increased, the frequency of “self” responses decreased compared to “delay” responses. All parameters and coefficients for the GLMM are detailed in Table 3.1. In contrast, the comparison between “other” and “delay” responses did not reveal a significant main effect of delay ($b = 0.003$, $SE = 0.001$, $t = 1.918$, $p = 0.055$, OR = 1.003). Only the intercept was significant ($b = -6.180$, $SE = 1.414$, $t = -4.370$, $p < 0.001$, OR = 0.002), indicating that “other” responses were infrequent in the delay condition. Detailed parameters and coefficients for the GLMM are provided in Table 3.2. Frequency distributions for fixation durations are shown in Figure 3.4C, and saccade amplitudes are presented in Figure 3.4D.

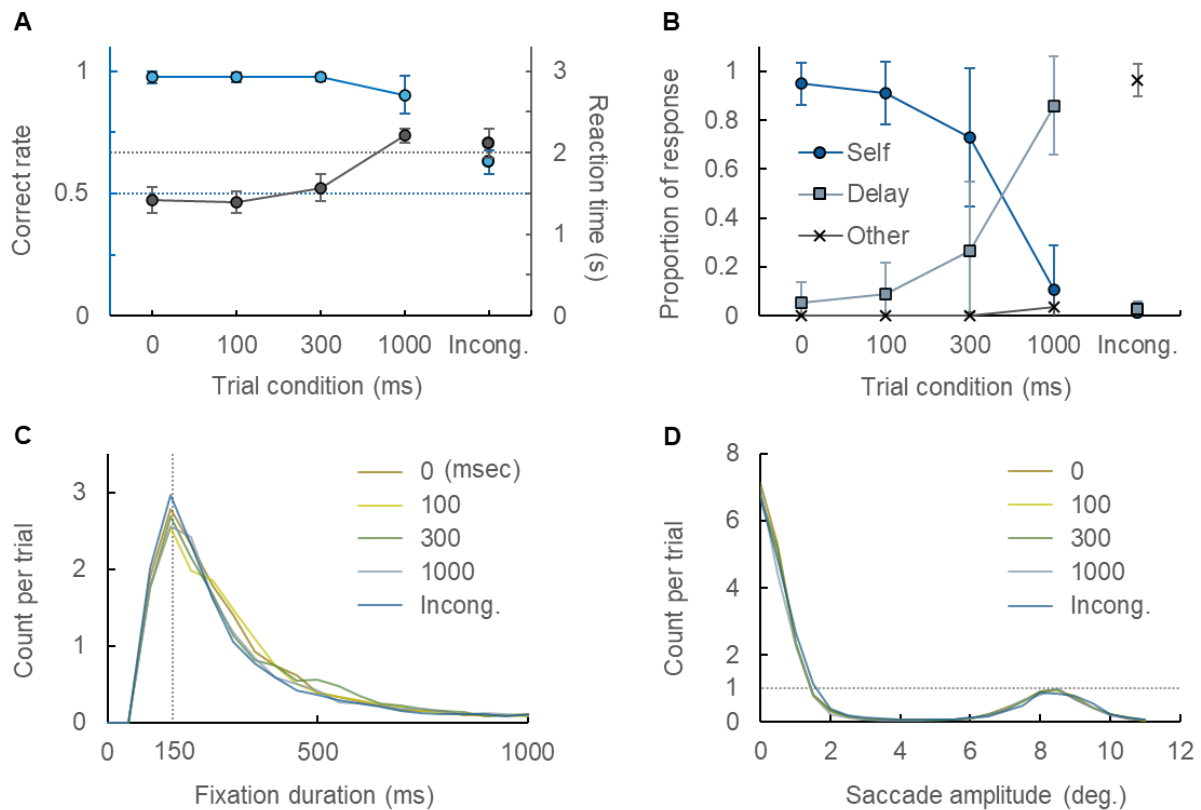


Figure 3.4. A: Correct rate (left y-axis) and reaction time (right y-axis) as functions of trial condition. The dashed blue line represents the chance level for the correct rate, while the grey dashed line marks the time when gaze-contingent stimuli were presented. B: The proportion of participants reporting authorship as self, delay, or other. For panels A and B means and standard deviations are calculated across participants ($N = 23$). C: Distribution of fixation durations across trial conditions, showing means for each condition. Representative values for each bin are positioned at the bin's minimum value (e.g., 100 ms for the 100–150 ms bin). The dashed line denotes a fixation duration of 150 ms. D: Distribution of saccade amplitudes across trial conditions, indicating means for each condition. Representative values for each bin are placed at the bin's minimum value (e.g., 1.5° for the $1.5\text{--}2^\circ$ bin). The dashed line represents a count of 1 per trial.

Table 3.1. Estimated covariance parameters and fixed coefficients for the authorship questionnaire, comparing “self” responses to “delay” responses.

Covariance parameter	Estimate	Standard Error	Z score	P value
Participants	1.625	0.573	2.836	0.005

Fixed effects	Coefficient	Standard Error	T score	P value
Intercept	3.268	0.308	10.602	0.000
Delay	-0.006	0.000	-19.982	0.000

Table 3.2. Estimated covariance parameters and fixed coefficients for the authorship questionnaire, comparing “other” responses to “delay” responses.

Covariance parameter	Estimate	Standard Error	Z score	P value
Participants	1.357	0.858	1.582	0.114

Fixed effects	Coefficient	Standard Error	T score	P value
Intercept	-6.180	1.414	-4.370	0.000
Delay	0.003	0.001	1.918	0.055

Delay-Modulated Activation

I examined the brain regions activated by the magnitude of the delay (in millisecond units). The right angular gyrus (Figure 3.5, Table 3.3) has been revealed as a statistically significant area. This area aligns with prior studies on limb movement [37].

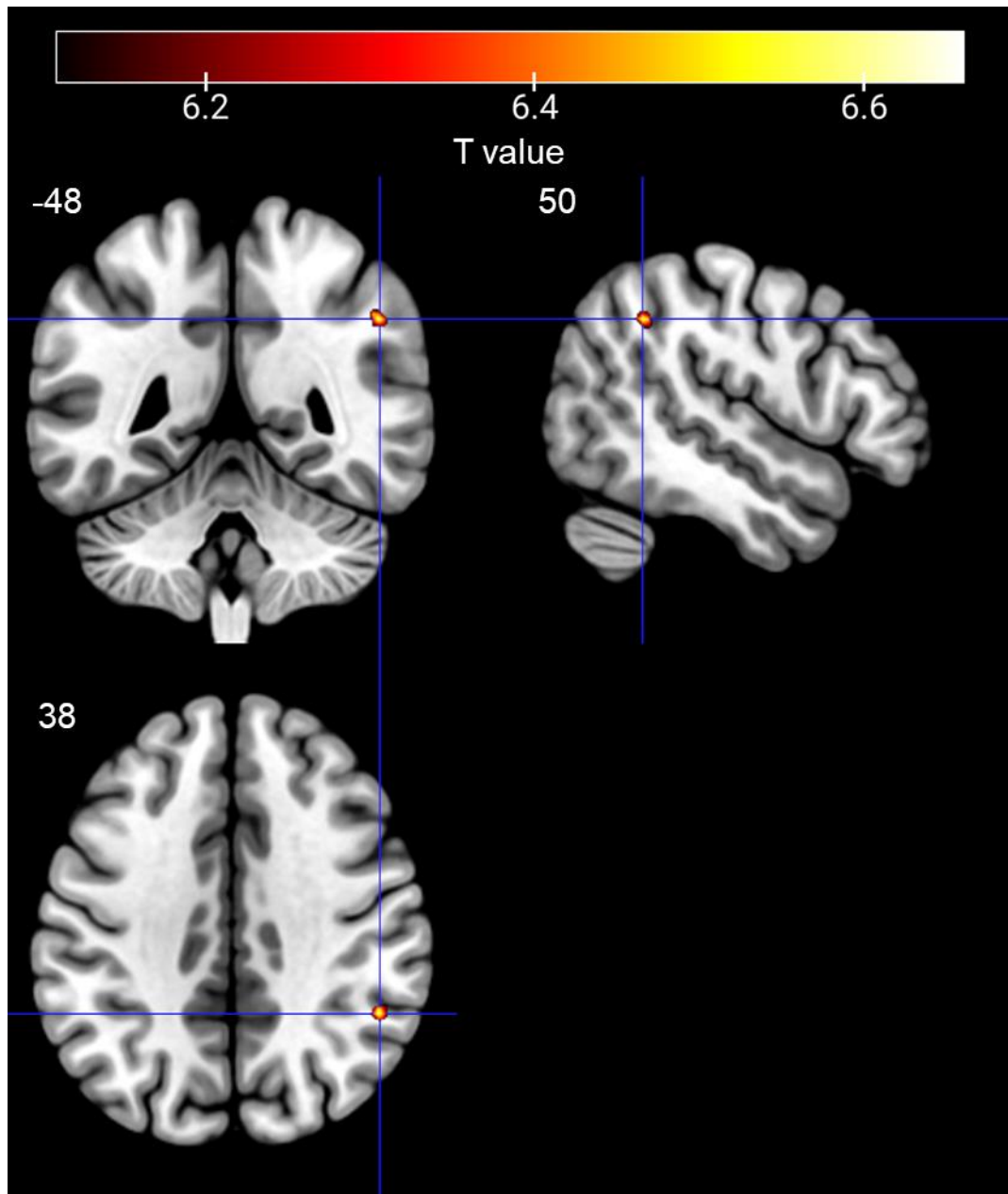


Figure 3.5. The delay-modulated activation was superimposed on an MNI template, an averaged T1 image from 152 individuals [115], using MRICroGL (<https://www.nitrc.org/projects/mricrogl/>) software. Significant delay-modulated activation was observed in the angular gyrus of the right hemisphere. The color bar indicates the range of t-values.

Table 3.3. Group analysis of delay-modulated brain region.

Cluster-level	MNI coordinate	Peak-level	Z-value	Hem.	Anatomical region
---------------	----------------	------------	---------	------	-------------------

Cluster size (mm ³)	FWE_p	x	y	z	FWE_p			
144	0.006	50	-48	38	0.018	4.88	R	Angular gyrus (Inferior parietal lobule, BA 40)

The threshold height of activation was $p < 0.05$ FWE corrected for multiple comparisons. x , y and z are stereotaxic coordinates (mm). The anatomical region in parentheses represents the label based on the Talairach atlas and Brodmann area—hem, Hemisphere; R, right.

Delay-Related Activation

The repeated one-way ANOVA in delayed conditions revealed relatively broad significant activation in the delay-modulated analysis at the right inferior parietal lobule (Figure 3.6, Table 3.4). Further, the left inferior occipital gyrus and right middle temporal gyrus were shown to be affected by the delay.

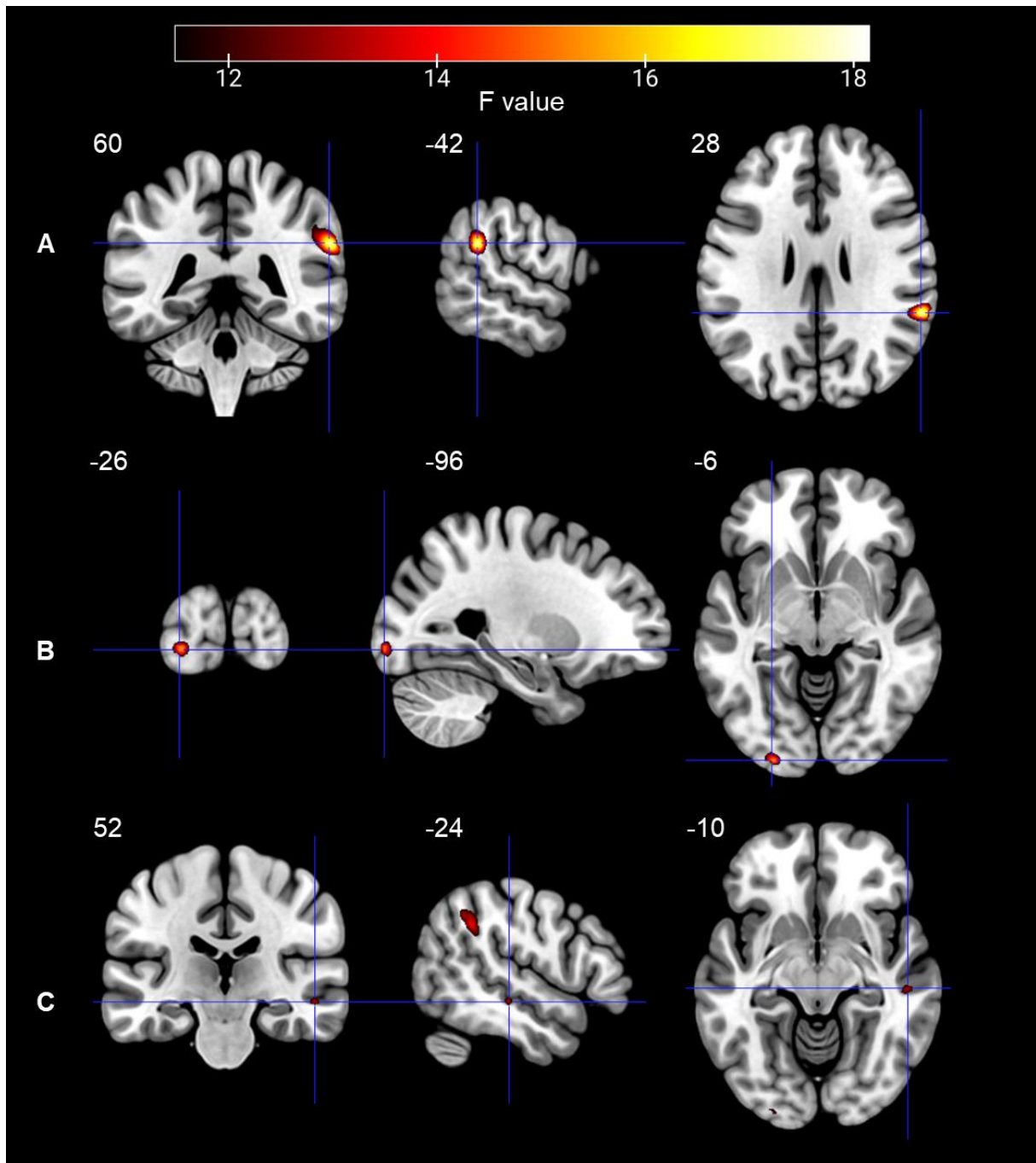


Figure 3.6. The delay-related activation was overlaid on an MNI template, an averaged T1 image derived from 152 individuals [115], using MRICroGL

(<https://www.nitrc.org/projects/mricrogl/>) software. Significant activation clusters were observed in the right supramarginal gyrus (A), left inferior occipital gyrus (B), and right middle temporal gyrus (C). The color bar indicates the range of f-values.

Table 3.4. Group analysis of delay-related brain region.

Cluster-level Cluster size (mm ³)	FWE_p	MNI coordinates			Peak- level	Z-value	Hem.	Anatomical region
		x	y	z	FWE_p			
1904	<0.0005	60	-42	28	<0.0005	5.82	R	Supramarginal gyrus (Inferior parietal lobule, BA 40)
		52	-46	34	0.006	5.10	R	Angular gyrus (Supramarginal gyrus, BA 40)
376	0.001	-26	-96	-6	0.001	5.41	L	Inferior occipital gyrus (Cuneus, BA 18)
96	0.010	52	-24	-10	0.007	5.06	R	Middle temporal gyrus (Superior temporal gyrus, BA 21)

The threshold height of activation was $p < 0.05$ FWE corrected for multiple comparisons. x, y and z are stereotaxic coordinates (mm). The anatomical region in parentheses represents the label based on the Talairach atlas and Brodmann area. Hem, Hemisphere; R, right; L, left.

To further investigate the direction of the delay's effect, average beta weights at the cluster located at the inferior parietal lobule were calculated from the 23 participants using the LOSO method. It is shown that the brain activation was increased when the delay was 1000 ms and also in the incongruity condition.

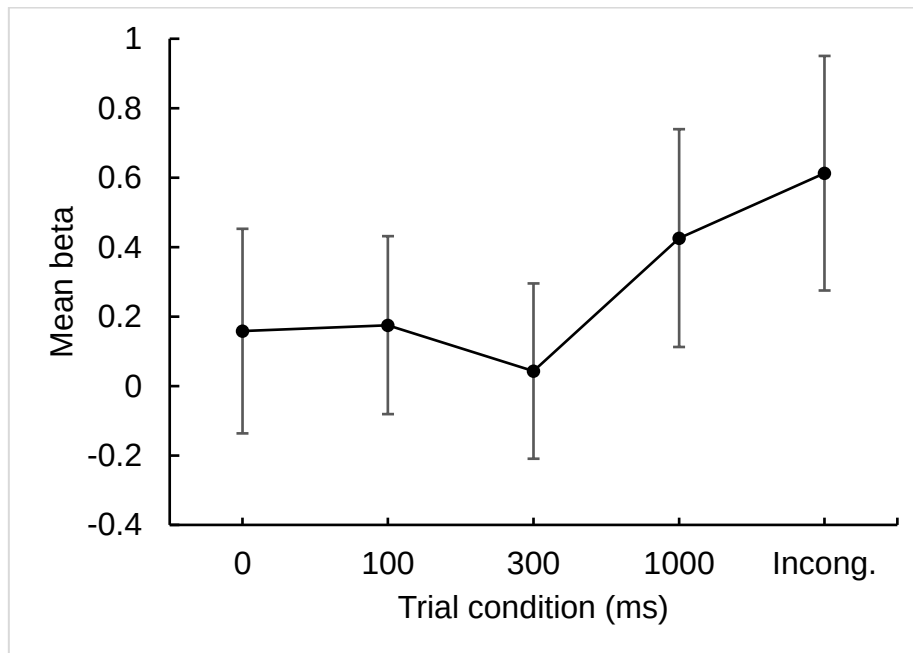


Figure 3.7. Means and standard deviations (error bars) were calculated for the averaged beta values within the manually created ROI for each condition across 23 participants. The delay-

related repeated ANOVA analysis identified the right supramarginal gyrus (MNI coordinates 60, -42, 28) as a significant activation region. The degree of activation was further investigated to determine the direction of the effect. Using the “Leave-One-Subject-Out” (LOSO) method [122], the mean beta values were calculated for each participant within the ROI, which was a manually selected cluster containing the MNI coordinates (60, -42, 28) from the significant clusters identified in the delay-related ANOVA analysis of the other 22 participants.

Incongruent vs. 0 ms Contrast Activation

Brain activation regions resulting from the contrast comparing “incong.” to “0 ms” condition revealed bilateral activations at the supramarginal gyrus, middle frontal gyrus, and precuneus.

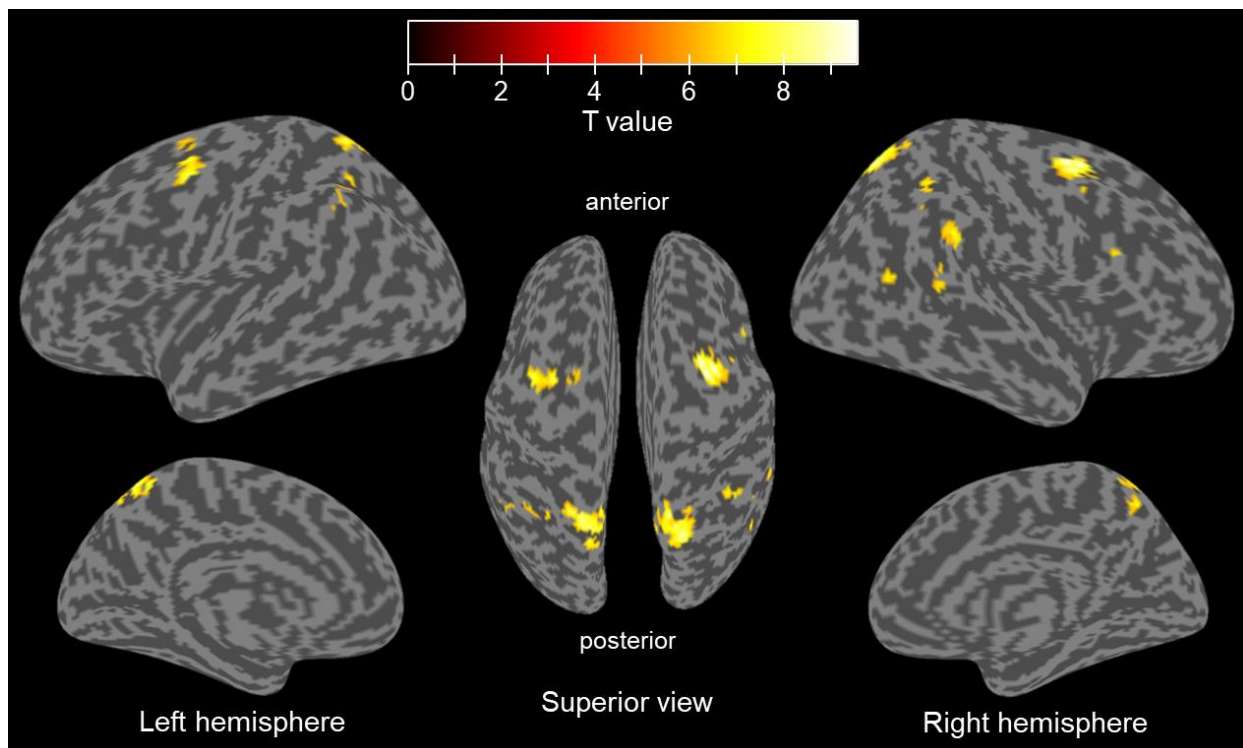


Figure 3.8. The activations comparing the incongruency to 0 ms conditions were overlaid on inflated canonical templates [Cortex_20484 surface mesh] using Statistical Parametric Mapping (SPM12, version 7771, FIL Methods Group) software. The color bar indicates the

range of t-values.

Table 3.5. Group analysis of contrast brain region comparing incongruency to 0 ms condition.

Cluster-level	Cluster size (mm ³)	FWE_p	MNI coordinates			Peak-level FWE_p	Z-value	Hem.	Anatomical region
			x	y	z				
3512	<0.0005		26	-4	48	<0.0005	6.48	R	Middle frontal gyrus (Middle frontal gyrus, BA 6)
			38	2	48	0.013	5.01	R	Middle frontal gyrus (Middle frontal gyrus, BA 6)
4560	<0.0005		20	-68	58	<0.0005	5.98	R	Superior parietal lobule (Precuneus, BA 7)
			12	-58	64	<0.0005	5.76	R	Superior parietal lobule (Precuneus, BA 7)
			18	-60	56	0.001	5.62	R	Superior parietal lobule (Precuneus, BA 7)
			14	-60	50	0.003	5.33	R	Precuneus (Precuneus, BA 7)
			10	-62	48	0.005	5.22	R	Precuneus (Precuneus, BA 7)
3704	<0.0005		-14	-58	52	<0.0005	5.83	L	Superior parietal lobule (Precuneus, BA 7)
			-16	-58	56	<0.0005	5.82	L	Superior parietal lobule (Precuneus, BA 7)
			-18	-58	64	0.001	5.58	L	Superior parietal lobule (Precuneus, BA 7)
			-10	-58	60	0.002	5.41	L	Superior parietal lobule (Precuneus, BA 7)
2784	<0.0005		-34	-4	42	0.001	5.61	L	Precentral gyrus (Middle frontal gyrus, BA 6)
			-24	-2	56	0.001	5.50	L	Superior frontal gyrus (Sub-Gyral, BA 6)
544	<0.0005		-58	-52	40	0.004	5.30	L	Supramarginal gyrus (Inferior Parietal Lobule, BA 40)
			-46	-58	52	0.020	4.91	L	Angular gyrus (Inferior Parietal Lobule, BA 40)
520	<0.0005		36	-40	38	0.004	5.30	R	Superior parietal lobule

								(Inferior Parietal Lobule, BA 40)
848	<0.0005	60	-42	32	0.006	5.20	R	Supramarginal gyrus (Supramarginal Gyrus, BA 40)
472	<0.0005	42	-58	12	0.007	5.17	R	Middle temporal gyrus (Superior Temporal Gyrus, BA 22)
424	<0.0005	-28	-54	46	0.007	5.14	L	Superior parietal lobule (Superior Parietal Lobule, BA 7)
		-36	-52	44	0.020	4.90	L	Superior parietal lobule (Inferior Parietal Lobule, BA 40)
208	0.002	46	16	32	0.008	5.13	R	Middle frontal gyrus (Middle Frontal Gyrus, BA 9)
288	0.001	52	-42	12	0.013	5.00	R	Superior temporal gyrus (Superior Temporal Gyrus, BA 13)
128	0.004	54	-46	46	0.028	4.83	R	Supramarginal gyrus (Inferior Parietal Lobule, BA 40)

The threshold height of activation was $p < 0.05$ FWE corrected for multiple comparisons. x, y and z are stereotaxic coordinates (mm). The anatomical region in parentheses represents the label based on the Talairach atlas and Brodmann area. Hem, Hemisphere; R, right; L, left.

Discussion

Behaviors

The correct rates were above the chance level and near 100% in the delayed conditions, indicating that participants conducted the visual search. Interestingly, the correct rates seemed above the chance level in the incongruency condition. This suggests that participants could still perform the visual search using their peripheral vision when the gaze-contingent window was located at the periphery. Reaction times showed that participants could respond within the time the gaze-contingent display was presented. Although their reaction times exceeded 2 s in the 1000 ms delayed condition and the incongruency condition, they did not surpass 3 s, which is the time when the questionnaire began to be

presented during the last trial. Therefore, I considered that there was no task-relevant brain activity during the baseline period set with the questionnaire time.

I now discuss the calculated fixation duration (Figure 3.4C) and saccade amplitude (Figure 3.4D) distributions. The distance from the central Chinese character to one of the surrounding four characters was 8.5 degrees. The distinct peak at 8–9 degrees in the saccade amplitude distribution represents the participants' saccades to move their gaze to one of the surrounding characters. Although the cumulative values of the saccade amplitude distribution do not confirm that participants made only a single saccade per trial, the distribution of fixation durations and saccade amplitudes indicates that the participants executed similar eye movements regardless of trials and conditions. Thus, I considered the participants' eye behaviors consistent with the instructions to execute one saccade and then maintain their gaze during the visual search.

In the categorical questionnaire, participants' response proportion of "self" maintained dominancy until the 300 ms delay. This was longer than Chapters 2 and 3. I assume the amount of internal sensory information, including the degree of freedom in action and HCI's feedback, leads to the difference. The current experiment restricts the visual search to only 2 seconds. Further, the participants were allowed to saccade to only one character among the four surrounding characters. Meanwhile, in Chapters 2 and 3, participants were allowed to visual search without time constraints and possibly to saccade any location among the nine characters. A relatively minor amount of sensory feedback was assumed to hinder the participant's delayed perception. Nonetheless, the results align with Chapters 2 and 3 in that the participant perceived the delay in the 1000 ms delayed condition, and they responded with "other" in only the uncontrollable condition. The uncontrollable condition served the role that led participants to perceive that the computer was the event author but not the self.

Brain activations

The delay-modulated analysis revealed the right angular gyrus as the significantly modulated brain area. As the participants' "delay" responses increased with the delay in the current eye-gaze HCI, I assert that this region is associated with the participant's perception of the delay. The averaged beta weights from the delay-related analysis showed that activation in the right supramarginal and angular gyri (rSMG and rAG) increased only at the 1000 ms and incongruency conditions. This implies that these areas are related to the participants' subjective perception of the delay rather than the delay itself. In other words, the right angular gyrus does not activate in proportion to the discrepancy between intended eye movements and the HCI's update. This result aligns with the previous study on limb movement [37] in that the earlier study also revealed that the angular gyrus activation was not proportional to the amount of the delay but increased when the participants responded to the perceived delay at the limb's controlled stimuli. Furthermore, this region is related to the delay perception and the perception of "I did not control the stimuli, but other agents such as the computer did." With the current delay-related analysis, a broader area of supramarginal and angular gyrus averaged beta weights were increased in both the 1000 ms delayed and incongruency conditions. Hence, the activation is also related to the participants' subjective judgment of the event author. Specifically, the activation is associated with judging the author to another (i.e., computer). In particular, activation is associated with a diminished sense of agency. Such results again align with the previous study in that this region was activated when the participant reported the stimuli were authored by another agent but not self [37]. Further alignments in the SMG and rAG can be confirmed with the previous studies involving hand and limb movement [124,125].

In the delay-related analysis, the middle temporal gyrus was revealed as significant. Fine parcellation of the middle temporal gyrus was done by Xu et al. [126]. Based on their parcellation, the activated area from the current study seems to overlap with the sulcus

middle temporal gyrus (sMTG; no exact coordinates given in the study [126]), which was introduced to be related to decoding another person's gaze direction and strongly connected with the supramarginal gyrus. Decoding another's gaze with sMTG was also shown with the social gazing experiment [127]. In this experiment, participants reported the direction of the head or eye gaze on the stimuli's human face. When participants were decoding the direction of the eye gaze, their brain activity in a region with MNI coordinates 53, -22, -12—close to the current result—showed decreased activation compared to both the baseline and the head-gaze decoding condition. The previous study labeled the location as the superior temporal sulcus. Accordingly, by inspection of the current results of the brain image, the activated region is also close to the superior temporal sulcus. For the participants, the perception of another agent (i.e., computer) moving a gaze-contingent window and another person's presence, including sensing their gaze, involves different information processing pathways. This distinction poses a challenge when discussing the current activations. Meanwhile, the superior temporal sulcus is suggested to represent the other agent [128], which might explain the current activations in that the participants perceived the computer as the other agent.

The inferior occipital gyrus was also significant in the delay-related analysis. Based on the Brodmann area, the activation was located in the secondary visual area. In the present study, the eye-gaze HCI's delay increased the time spent inspecting the blurred character in the participants' central vision at the newly fixated gaze location. I speculate that this increased inspection time might be related to the activation in the inferior occipital gyrus.

The "incongruency—1000 ms" contrast revealed bilateral activation at the supramarginal gyrus, middle frontal gyrus, and precuneus. Regarding the middle frontal gyrus and precuneus (precentral gyrus), I suggest that these activations were primarily governed by the attention system, as previously suggested. [129,130]. The attention system shows increased activity when the participants detect the appearance of salient targets at an

unexpected location. This increased activity aligns with the activation demonstrated by the “incongruency —1000ms” contrast. Since I instructed the participants to visually search their periphery when the window was outside their central vision, they would likely attend to their periphery in the incongruency condition. The high-resolution character, with increased brightness compared to the blurred character, likely attracted the participants’ attention in the periphery during the incongruency condition. Bilateral activation at the supramarginal gyrus again supports judging the “other” agent (i.e., computer) as the event author. Similar results were reported in the study by Ohata et al. [125]. The study instructed participants to manipulate a joystick to move a cursor. They were further instructed to trace a presented sinusoidal target path at a specified pace (i.e., 2 seconds per cycle). During the trial, a pre-recorded trajectory was mixed with the participant’s input to generate a spatial discrepancy between the input and the updated cursor position. The brain activations related to the joystick-cursor distance revealed significant bilateral activations in the precentral gyrus and unilateral (left hemisphere) activation in the inferior parietal lobe. The participants likely attended to the cursor’s appearance in unexpected locations, which might have led to similar activations.

Conclusion

Chapter 4 utilized fMRI to examine the neural mechanisms underlying the sense of agency in gaze-contingent HCI. The current findings identified the right angular gyrus (rAG) as a critical brain region involved in disruptions in perception caused by temporal discrepancies between eye movements and corresponding interface updates. This neural activation parallels previous findings in hand or limb movements, suggesting that the internal signals related to eye movements are processed similarly in perceiving action author.

Chapter 5

General Discussion

Based on Chapter 2 and 3, shift in strategies and mode transitions were observed in response to delays, these findings suggest that future applications of eye-gaze HCI should prioritize designs that preserve the first-person view and ensure seamless interaction. The “wait-and-read” strategy, which embodies real-time perceptual control, is crucial for maintaining an immersive eye-gaze experience. In this mode, users can monitor visual feedback and adjust their actions in real-time, keeping a direct connection between eye movements and interface responses. However, as delays become more pronounced, this immediate interaction begins to break down, leading users to adopt a “pre-registering” strategy, anticipating delays and planning their gaze shifts. This shift from perceptual agency to belief-based control disrupts the intuitive, spontaneous interaction expected in effective eye-gaze systems. Therefore, future designs must minimize delays to maintain the immediacy and fluidity for a genuinely immersive first-person experience in eye-gaze applications.

In some articles, in addition to the sense of agency, the sense of ownership—the sensation that our body parts, feelings, or thoughts belong to the self—is acknowledged. Since the current study mainly focused on the relationship between eye movement and its feedback, this sensation was not proactively examined. Possibly, the visual version of the sense of ownership can be close to the sensation of first-person view-ness. Since many researchers have pointed out that the number of first-person view studies along with the self-body perception, such as the present study, is quite limited, discussing the visual sense of ownership with today’s scientific findings is challenging. Further research may be required.

This study demonstrated that the internal representation of non-self (or other) versus self in the right angular gyrus (rAG) played a crucial role in the sense of agency. This brain area is known for its importance in social cognition. These results provide important

implications for understanding the integration of social aspects in the brain and their impact on the sense of agency in eye-gaze research contexts. Then, the question arises: Why is the so-called social brain involved in the visual information process? One possibility is that in comparator models of the visual stabilities [12,13], the brain distinguishes between self-generated and non-self-generated retinal image displacement before and after eye movements to achieve visual stability and object motion perception. This ability allows the visual system to cancel out self-related signals, indicating that the processing of self-related information process is involved. How oculomotor signals and visual information are integrated into the social brain remains an area for further exploration.

Limitations

While the experiments provided insights, several limitations must be acknowledged. The sample size was relatively small, primarily due to the resource-intensive nature and the extended duration of the experimental sessions. The participant pool is constrained to a specific age bracket and the use of a particular distribution. This may limit the generalizability of the findings. Additionally, the controlled laboratory settings of the experiments may not fully capture the complexities of real-world HCI scenarios, where factors such as user fatigue, environmental distractions (e.g., network), and varying levels of task engagement can influence the sense of agency. The use of subjective reports to measure the sense of agency, while informative, is often prone to systematic bias. Future research could potentially gain from incorporating objective measures through signal detection frameworks.

Future Directions

Future research would also address the limitations identified and build upon the current findings. Increasing the sample size and diversifying the participant pool can enhance the generalizability of the results. Additionally, extending the experimental paradigms with

various settings, such as real-world applications of eye-gaze HCI, can provide a deeper understanding of how temporal delays and variability impact the sense of agency in everyday use.

Investigating the long-term effects of delayed feedback on user behavior and the sense of agency can provide insights into how users adapt over time. Moreover, exploring the interplay between different sensory modalities, such as combining eye gaze with hand movements in HCI, can offer a more comprehensive understanding of multimodal interactions and their impact on the sense of agency.

Advancements in neuroimaging techniques, such as high-resolution fMRI and real-time neurofeedback, can further elucidate the neural mechanisms underlying the sense of agency. Examining how brain regions interact during delayed and jittery feedback scenarios can provide a more detailed neural map of agency-related processes.

The question remains whether the oculomotor signals were directly accessed via consciousness. From Taylor's [86] suggestion, the shift signals in the participants' attention might be the root cause underlying the corollary discharge and sense of agency. There are overlaps with the current findings that the involvement of the parietal lobe (superior parietal lobule or inferior parietal sulcus, SPL/IPS) (Experiment 3) and the connection to attention distribution (Experiments 1 and 2) might be linked to the corollary discharge mechanisms discussed by Taylor [95]. Further research could explore the detailed interaction between these regions, mainly focusing on how attention control circuits in the SPL/IPS and rAG contribute to the sense of agency.

Separating eye movements and attended location would be a way to investigate the interaction between oculomotor signals and attention shift signals in the sense of agency. However, controlling HCI with the participants' attended location is a challenging task. Although several studies have attempted to do so, with steady-state visual-evoked potential

(SSVEP), signals from electroencephalography (EEG) electrodes could not distinguish between overt and covert attention [131,132]. Limited resources in attention distribution have been suggested as a limiting factor in the experiment paradigm [132]. Inducing unintentional oculomotor signals might lead participants to maintain the attended location while their eyes move. Creating a baseline condition, for example, by employing transcranial magnetic stimulation to provoke unintentional actions [16,133], is crucial for a comprehensive investigation of this issue. This approach, however, represents a significant challenge for future research.

Exploring intentional binding through paradigms such as the time interval estimates [134] or the Libet clock [16] could offer valuable insights into the scope of this dissertation. However, the chronostasis effect—where perceived time expands immediately after a saccade—complicates experimental design [66]. Moreover, intentional binding paradigms are not suitable for studying jittery delays within a continuous action-effects system. Nevertheless, examining brain activations during intentional binding experiments could substantially enhance our comprehension of the phenomena under investigation in this dissertation.

Notably, further studies involving angular-biased visual feedback would be interesting for investigating the misattribution of an action author to another agent, as in the studies by Farrer et al. [19,67]. Moreover, placing more characters in the array and decreasing the distance among the characters would be attractive for determining whether task difficulty influences the establishment of authorship, as previous studies examined the relationship between task performance and sense of agency [22,24]. Modulation in task difficulty and temporal delay might reveal the participants' limited attention distribution, thus leading to insights.

Based on the study results, an eye-gaze HCI system would be a feasible application

that detects a user's altering sense of agency by monitoring their eye movements or EEG signals (in the rAG). Considering that the sense of agency attenuates when the display update does not match the user's own prediction, the system's alteration detection also suggests the user's inability to adapt to the temporal mismatch in the display update. By measuring this inability to adjust, the system could manipulate the discrepancy (i.e., delay) to a degree where the user perceives the display update at their adaptation threshold. Thus, the system can balance between processing cost and user performance using the eye-gaze interface.

If a prediction signal in eye-gaze HCI could be measured, further applications could measure the user's intention to manipulate the interface, thus allowing the interface to distinguish between the user's observational and interface-operational eye movements. This capability would contribute significantly to enhancing computer accessibility by increasing the overall modalities of computer input.

Conclusion

The subjective reports and the neural evidence in three experiments support the notion that the eye can be perceived as a causal initiator in HCI, further expanding the understanding of the sense of agency from hand and limb movements to include eye movements. Additionally, the current investigations differentiate from the previous studies with limb movements in that the action refinement was systematically investigated, shedding light on a new approach to examining the sense of agency. Integrating these findings, the unique impact of action-feedback mismatch in eye movements influences attention deployment and action refinement, offering new approaches to understanding the sense of agency distinct from previous research. The present study significantly advances the development guidelines of eye-gaze interfaces in environments with delays and variability. The insights provided by these studies lay the groundwork for designing more effective and resilient delay-tolerant interfaces, enhancing user experience in diverse HCI contexts. This dissertation bridges

theoretical and practical gaps in understanding the sense of agency in gaze-contingent HCI, contributing valuable knowledge and informing future research and development in this domain.

Acknowledgements

I would like to express my profound gratitude to Professor Takako Yoshida, who has been a guiding light throughout my research journey since my master's studies. Her thoughtful insights and unwavering encouragement have been a constant source of support whenever I sought guidance, playing an essential role in shaping the direction of my academic path. I am especially thankful for her invaluable assistance in interpreting research feedback, whether from my own work or journal comments. Her ability to help me grasp the underlying research directions ensured that I stayed focused and on course.

I extend my sincere appreciation to Professor Tohru Yagi, who served as my co-advisor during my master's. His invaluable advice on the mindset required as a researcher helped me build a strong foundation that has guided me throughout my academic journey.

I am deeply grateful to Professor Hirohiko Kaneko, my co-advisor during my Ph.D., for his patient and thoughtful guidance. His continuous support was instrumental in ensuring the stability and progress of my research throughout the doctoral process.

Special thanks go to Professors Hiroyuki Akama and Yasuharu Koike, whose critical insights since my master's have played a key role in broadening my perspectives. Their guidance helped me uncover important aspects of my research that I had not previously considered.

I would also like to express my gratitude to Professor Yasuyoshi Yokokohji from Kobe University, who has been an essential guide throughout my research. His direction from an engineering perspective consistently provided me with valuable insights into the trajectory of my work.

Lastly, I would like to thank Professors Takehiro Nagai and Natsue Yoshimura. Their insightful comments greatly improved the rigor of my work and were instrumental in strengthening the validity of my experiments.

I would like to extend my heartfelt thanks to Maki Kawasaki, who has not only managed all the administrative work in the lab but also cared for me personally throughout both my master's and doctoral studies. Her kindness and attention, akin to the care of a parent, have been deeply appreciated, and I am forever grateful for her unwavering support.

I am also grateful to the professors, staff, and colleagues at the Tokyo Tech Academy for Leadership. Their guidance and encouragement, especially in areas beyond research, have been vital to my growth as both a scholar and an individual.

I would like to acknowledge the generous financial support I received from the Japan Science and Technology Agency (JST)'s Support for Pioneering Research Initiated by the Next Generation (SPRING) Program. Their funding provided crucial stability throughout my Ph.D., allowing me to focus fully on my research and pursue ambitious goals.

Additionally, I am thankful for the research funding from the ROBOT Industrial Basic Technology Collaborative Innovation Partnership grant provided by Japan's New Energy and Industrial Technology Development Organization (NEDO). This funding, though part of my professor's research grant, was vital in advancing my studies and publishing my work, enabling me to bring my research ambitions to life.

I would like to thank Eric Redlinger, a former Ph.D. colleague, and Yusuke Hachisuga, my colleague during my master's studies, for their valuable feedback and discussions.

Finally, my deepest and most heartfelt gratitude goes to my family. Without their unwavering belief in me, I would not have had the courage to embark on this journey, let alone see it through. Their financial support during the most difficult times made it possible for me to complete my Ph.D., but more importantly, their emotional strength, love, and encouragement sustained me every step of the way. I owe everything to my family, and I would not be where I am today without them.

References

1. Gallagher S. Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn Sci.* 2000;4: 14–21. doi:10.1016/S1364-6613(99)01417-5
2. Newen A, Vogeley K. Self-representation: Searching for a neural signature of self-consciousness. *Conscious Cogn.* 2003;12: 529–543. doi:10.1016/S1053-8100(03)00080-1
3. Kircher TTJ, Leube DT. Self-consciousness, self-agency, and schizophrenia. *Conscious Cogn.* 2003;12: 656–669. doi:10.1016/S1053-8100(03)00071-0
4. Weiskrantz L, Elliott J, Darlington C. Preliminary Observations on Tickling Oneself. *Nature.* 1971;230: 598–599. doi:10.1038/230598a0
5. Claxton G. Why Can't We Tickle Ourselves? *Percept Mot Skills.* 1975;41: 335–338. doi:10.2466/pms.1975.41.1.335
6. Blakemore S-J, Wolpert DM, Frith CD. Central cancellation of self-produced tickle sensation. *Nat Neurosci.* 1998;1: 635–640. doi:10.1038/2870
7. Blakemore S-J, Wolpert D, Frith C. Why can't you tickle yourself? *Neuroreport.* 2000;11: R11–R16. doi:10.1097/00001756-200008030-00002
8. Blakemore S-J, Frith CD, Wolpert DM. Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *J Cogn Neurosci.* 1999;11: 551–559. doi:10.1162/089892999563607
9. von Holst E. Relations between the central Nervous System and the peripheral organs. *The British Journal of Animal Behaviour.* 1954;2: 89–94. doi:10.1016/S0950-5601(54)80044-X
10. Cardoso-Leite P, Mamassian P, Schütz-Bosbach S, Waszak F. A New Look at Sensory Attenuation: Action-Effect Anticipation Affects Sensitivity, Not

- Response Bias. *Psychol Sci.* 2010;21: 1740–1745.
doi:10.1177/0956797610389187
11. Mifsud NG, Beesley T, Watson TL, Elijah RB, Sharp TS, Whitford TJ. Attenuation of visual evoked responses to hand and saccade-initiated flashes. *Cognition.* 2018;179: 14–22. doi:10.1016/j.cognition.2018.06.005
 12. Bridgeman B. A review of the role of efference copy in sensory and oculomotor control systems. *Ann Biomed Eng.* 1995;23: 409–422.
doi:10.1007/BF02584441
 13. Bridgeman B. How the Brain Makes the World Appear Stable. *Iperception.* 2010;1: 69–72. doi:10.1068/i0387
 14. Haggard P. Sense of agency in the human brain. *Nat Rev Neurosci.* 2017;18: 196–207. doi:10.1038/nrn.2017.14
 15. Frith CD, Blakemore S-J, Wolpert DM. Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Res Rev.* 2000;31: 357–363. doi:10.1016/S0165-0173(99)00052-1
 16. Haggard P, Clark S, Kalogeras J. Voluntary action and conscious awareness. *Nat Neurosci.* 2002;5: 382–385. doi:10.1038/nn827
 17. Kalckert A, Ehrsson H. Moving a Rubber Hand that Feels Like Your Own: A Dissociation of Ownership and Agency. *Front Hum Neurosci.* 2012;6.
doi:10.3389/fnhum.2012.00040
 18. Sato A, Yasuda A. Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition.* 2005;94: 241–255.
doi:10.1016/j.cognition.2004.04.003

19. Farrer C, Bouchereau M, Jeannerod M, Franck N. Effect of Distorted Visual Feedback on the Sense of Agency. *Behavioural Neurology*. 2008;19: 53–57. doi:10.1155/2008/425267
20. Farrer C, Valentin G, Hupé JM. The time windows of the sense of agency. *Conscious Cogn*. 2013;22: 1431–1441. doi:10.1016/j.concog.2013.09.010
21. Ebert JP, Wegner DM. Time warp: Authorship shapes the perceived timing of actions and events. *Conscious Cogn*. 2010;19: 481–489. doi:10.1016/j.concog.2009.10.002
22. Wen W, Yamashita A, Asama H. The Sense of Agency during Continuous Action: Performance Is More Important than Action-Feedback Association. *PLoS One*. 2015;10: e0125226-. doi:10.1371/journal.pone.0125226
23. Wen W, Yamashita A, Asama H. The influence of action-outcome delay and arousal on sense of agency and the intentional binding effect. *Conscious Cogn*. 2015;36: 87–95. doi:10.1016/j.concog.2015.06.004
24. Shibuya S, Unenaka S, Ohki Y. The Relationship Between the Virtual Hand Illusion and Motor Performance. *Front Psychol*. 2018;9. doi:10.3389/fpsyg.2018.02242
25. Moore JW, Obhi SS. Intentional binding and the sense of agency: A review. *Conscious Cogn*. 2012;21: 546–561. doi:10.1016/j.concog.2011.12.002
26. Synofzik M, Vosgerau G, Newen A. Beyond the comparator model: A multifactorial two-step account of agency. *Conscious Cogn*. 2008;17: 219–239. doi:10.1016/j.concog.2007.03.010
27. Wegner DM, Sparrow B, Winerman L. Vicarious Agency: Experiencing Control Over the Movements of Others. *J Pers Soc Psychol*. 2004;86: 838–848. doi:10.1037/0022-3514.86.6.838

28. Pfeiffer U, Schilbach L, Timmermans B, Jording M, Bente G, Vogeley K. Eyes on the Mind: Investigating the Influence of Gaze Dynamics on the Perception of Others in Real-Time Social Interaction. *Front Psychol.* 2012;3. doi:10.3389/fpsyg.2012.00537
29. Recht S, Grynszpan O. The sense of social agency in gaze leading. *Journal on Multimodal User Interfaces.* 2019;13: 19–30. doi:10.1007/s12193-018-0286-y
30. Brandi M-L, Kaifel D, Lahnakoski JM, Schilbach L. A naturalistic paradigm simulating gaze-based social interactions for the investigation of social agency. *Behav Res Methods.* 2020;52: 1044–1055. doi:10.3758/s13428-019-01299-x
31. Grynszpan O, Simonin J, Martin J-C, Nadel J. Investigating social gaze as an action-perception online performance. *Front Hum Neurosci.* 2012;6. doi:10.3389/fnhum.2012.00094
32. Gregori Grgič R, Crespi SA, de'Sperati C. Assessing Self-Awareness through Gaze Agency. *PLoS One.* 2016;11: e0164682-. doi:10.1371/journal.pone.0164682
33. Gutzeit J, Weller L, Muth F, Kürten J, Huestegge L. Eye did this! Sense of agency in eye movements. *Acta Psychol (Amst).* 2024;243: 104121. doi:10.1016/j.actpsy.2023.104121
34. Wen W. Does delay in feedback diminish sense of agency? A review. *Conscious Cogn.* 2019;73: 102759. doi:10.1016/j.concog.2019.05.007
35. Chambon V, Wenke D, Fleming SM, Prinz W, Haggard P. An Online Neural Substrate for Sense of Agency. *Cerebral Cortex.* 2013;23: 1031–1037. doi:10.1093/cercor/bhs059

36. Farrer C, Frith CD. Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *Neuroimage*. 2002;15: 596–603. doi:10.1006/nimg.2001.1009
37. Farrer C, Frey SH, Van Horn JD, Tunik E, Turk D, Inati S, et al. The Angular Gyrus Computes Action Awareness Representations. *Cerebral Cortex*. 2008;18: 254–261. doi:10.1093/cercor/bhm050
38. Sperduti M, Delaveau P, Fossati P, Nadel J. Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. *Brain Struct Funct*. 2011;216: 151–157. doi:10.1007/s00429-010-0298-1
39. Kesner L, Grygarová D, Fajnerová I, Lukavský J, Nekovářová T, Tintěra J, et al. Perception of direct vs. averted gaze in portrait paintings: An fMRI and eye-tracking study. *Brain Cogn*. 2018;125: 88–99. doi:10.1016/j.bandc.2018.06.004
40. Schilbach L, Eickhoff SB, Cieslik E, Shah NJ, Fink GR, Vogeley K. Eyes on me: an fMRI study of the effects of social gaze on action control. *Soc Cogn Affect Neurosci*. 2011;6: 393–403. doi:10.1093/scan/nsq067
41. Moore JW. What Is the Sense of Agency and Why Does it Matter? *Front Psychol*. 2016;7. doi:10.3389/fpsyg.2016.01272
42. Sharma R, Pavlovic VI, Huang TS. Toward multimodal human-computer interface. *Proceedings of the IEEE*. 1998;86: 853–869. doi:10.1109/5.664275
43. Tobii. Tobii Dynavox. [cited 7 Sep 2023]. Available: <https://www.tobiidynavox.com/>
44. Microsoft. Get started with eye control in Windows. Retrieved from Microsoft Support. [cited 7 Sep 2023]. Available: <https://support.microsoft.com/en-us/windows/get-started-with-eye-control-in-windows-1a170a20-1083-2452-8f42-17a7d4fe89a9>

45. Anvari M, Broderick T, Stein H, Chapman T, Ghodoussi M, Birch DW, et al. The impact of latency on surgical precision and task completion during robotic-assisted remote telepresence surgery. *Computer Aided Surgery*. 2005;10: 93–99. doi:10.3109/10929080500228654
46. Gettys J, Nichols K. Bufferbloat: Dark Buffers in the Internet. *Commun ACM*. 2012;55: 57–65. doi:10.1145/2063176.2063196
47. Roy A, Pachua JL, Saha AK. An overview of queuing delay and various delay based algorithms in networks. *Computing*. 2021;103: 2361–2399. doi:10.1007/s00607-021-00973-3
48. Panzirsch M, Singh H, Krüger T, Ott C, Albu-Schäffer A. Safe Interactions and Kinesthetic Feedback in High Performance Earth-To-Moon Teleoperation. 2020 IEEE Aerospace Conference. 2020. pp. 1–10. doi:10.1109/AERO47225.2020.9172665
49. Imaida T, Yokokohji Y, Doi T, Oda M, Yoshikawa T. Ground-space bilateral teleoperation of ETS-VII robot arm by direct bilateral coupling under 7-s time delay condition. *IEEE Transactions on Robotics and Automation*. 2004;20: 499–511. doi:10.1109/TRA.2004.825271
50. Claypool M, Tanner J. The effects of jitter on the peceptual quality of video. *Proceedings of the Seventh ACM International Conference on Multimedia (Part 2)*. New York, NY, USA: Association for Computing Machinery; 1999. pp. 115–118. doi:10.1145/319878.319909
51. Gulliver SR, Ghinea G. The Perceptual and Attentive Impact of Delay and Jitter in Multimedia Delivery. *IEEE Transactions on Broadcasting*. 2007;53: 449–458. doi:10.1109/TBC.2007.896955

52. Hua J, Cui Y, Yang Y, Li H. Analysis and prediction of jitter of internet one-way time-delay for teleoperation systems. 2013 11th IEEE International Conference on Industrial Informatics (INDIN). 2013. pp. 612–617.
doi:10.1109/INDIN.2013.6622954
53. Kim S, Lee JY, Sung D. Shifted gamma distribution model for long-range dependent Internet traffic. *Communications Letters, IEEE*. 2003;7: 124–126.
doi:10.1109/LCOMM.2002.808400
54. Oboe R, Fiorini P. Issues on Internet-Based Teleoperation. *IFAC Proceedings Volumes*. 1997;30: 591–597. doi:10.1016/S1474-6670(17)44322-9
55. Sukhov AM, Astrakhantseva MA, Pervitsky AK, Boldyrev SS, Bukatov AA. Generating a function for network delay. *Journal of High Speed Networks*. 2016;22: 321–333. doi:10.3233/JHS-160552
56. Zhang W, He J. Modeling End-to-End Delay Using Pareto Distribution. *Second International Conference on Internet Monitoring and Protection (ICIMP 2007)*. 2007. p. 21. doi:10.1109/ICIMP.2007.26
57. Ieiri S, Hashizume M. Development of surgical assisted system for remote medicine. *Med Biol Eng*. 2011;49: 673–677. doi:10.11239/jsmbe.49.673
58. Sheridan TB. Space teleoperation through time delay: review and prognosis. *IEEE Transactions on Robotics and Automation*. 1993;9: 592–606.
doi:10.1109/70.258052
59. Chen H, Liu Z, Huang P, Kuang Z. Time-Delay Modeling and Simulation for Relay Communication-Based Space Telerobot System. *IEEE Trans Syst Man Cybern Syst*. 2022;52: 4211–4222. doi:10.1109/TSMC.2021.3090806
60. Quax P, Monsieurs P, Lamotte W, De Vleeschauwer D, Degrande N. Objective and subjective evaluation of the influence of small amounts of delay and jitter

- on a recent first person shooter game. *Proceedings of 3rd ACM SIGCOMM Workshop on Network and System Support for Games*. New York, NY, USA: Association for Computing Machinery; 2004. pp. 152–156.
doi:10.1145/1016540.1016557
61. Kelso JAS. On the Self-Organizing Origins of Agency. *Trends Cogn Sci*. 2016;20: 490–499. doi:10.1016/j.tics.2016.04.004
 62. Braun N, Debener S, Spsychala N, Bongartz E, Sörös P, Müller HHO, et al. The Senses of Agency and Ownership: A Review. *Front Psychol*. 2018;9. doi:10.3389/fpsyg.2018.00535
 63. Sperry RW. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol*. 1950;43: 482–489. doi:10.1037/h0055479
 64. Cavanaugh J, Berman RA, Joiner WM, Wurtz RH. Saccadic Corollary Discharge Underlies Stable Visual Perception. *The Journal of Neuroscience*. 2016;36: 31–42. doi:10.1523/JNEUROSCI.2054-15.2016
 65. Lindner A, Thier P, Kircher TTJ, Haarmerier T, Leube DT. Disorders of Agency in Schizophrenia Correlate with an Inability to Compensate for the Sensory Consequences of Actions. *Current Biology*. 2005;15: 1119–1124. doi:10.1016/j.cub.2005.05.049
 66. Yarrow K, Haggard P, Heal R, Brown P, Rothwell JC. Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*. 2001;414: 302–305. doi:10.1038/35104551
 67. Farrer C, Franck N, Georgieff N, Frith CD, Decety J, Jeannerod M. Modulating the experience of agency: a positron emission tomography study. *Neuroimage*. 2003;18: 324–333. doi:10.1016/S1053-8119(02)00041-1

68. Albert R, Patney A, Luebke D, Kim J. Latency Requirements for Foveated Rendering in Virtual Reality. *ACM Trans Appl Percept.* 2017;14: 1–13.
doi:10.1145/3127589
69. Loschky LC, Wolverton GS. How Late Can You Update Gaze-Contingent Multiresolutional Displays without Detection? *ACM Trans Multimedia Comput Commun Appl.* 2007;3. doi:10.1145/1314303.1314310
70. Geisler WS, Perry JS. Real-time foveated multiresolution system for low-bandwidth video communication. *ProcSPIE.* 1998. pp. 294–305.
doi:10.1117/12.320120
71. Tobii. Tobii spotlight Technology™. [cited 7 Sep 2023]. Available:
<https://www.tobii.com/resource-center/tech-talks/tobii-spotlight-technology-benchmark-test-on-vive-pro-eye>
72. Guenter B, Finch M, Drucker S, Tan D, Snyder J. Foveated 3D Graphics. *ACM Trans Graph.* 2012;31. doi:10.1145/2366145.2366183
73. Stengel M, Grogorick S, Eisemann M, Magnor M. Adaptive Image-Space Sampling for Gaze-Contingent Real-time Rendering. *Computer Graphics Forum.* 2016;35: 129–139. doi:10.1111/cgf.12956
74. Mohanto B, Islam ABMT, Gobbetti E, Staadt O. An integrative view of foveated rendering. *Comput Graph.* 2022;102: 474–501.
doi:10.1016/j.cag.2021.10.010
75. Anstis SM. A chart demonstrating variations in acuity with retinal position. *Vision Res.* 1974;14: 589–592. doi:10.1016/0042-6989(74)90049-2
76. Reingold EM, Loschky LC, McConkie GW, Stampe DM. Gaze-Contingent Multiresolutional Displays: An Integrative Review. *Hum Factors.* 2003;45: 307–328. doi:10.1518/hfes.45.2.307.27235

77. Patney A, Salvi M, Kim J, Kaplanyan A, Wyman C, Benty N, et al. Towards Foveated Rendering for Gaze-Tracked Virtual Reality. *ACM Trans Graph.* 2016;35. doi:10.1145/2980179.2980246
78. Cajar A, Schneeweiß P, Engbert R, Laubrock J. Coupling of attention and saccades when viewing scenes with central and peripheral degradation. *J Vis.* 2016;16: 8. doi:10.1167/16.2.8
79. Laubrock J, Cajar A, Engbert R. Control of fixation duration during scene viewing by interaction of foveal and peripheral processing. *J Vis.* 2013;13: 11–11. doi:10.1167/13.12.11
80. Cajar A, Engbert R, Laubrock J. Spatial frequency processing in the central and peripheral visual field during scene viewing. *Vision Res.* 2016;127: 186–197. doi:10.1016/j.visres.2016.05.008
81. Lakens D. Sample Size Justification. *Collabra Psychol.* 2022;8: 33267. doi:10.1525/collabra.33267
82. Foulsham T, Teszka R, Kingstone A. Saccade control in natural images is shaped by the information visible at fixation: evidence from asymmetric gaze-contingent windows. *Atten Percept Psychophys.* 2011;73: 266–283. doi:10.3758/s13414-010-0014-5
83. Reingold EM, Loschky LC. Saliency of peripheral targets in gaze-contingent multiresolutional displays. *Behavior Research Methods, Instruments, & Computers.* 2002;34: 491–499. doi:10.3758/BF03195478
84. Kim J, Yoshida T. Sense of agency at a temporally-delayed gaze-contingent display. *PLoS One.* 2024. doi:10.1371/journal.pone.0309998
85. Castelhana MS. Eye movements during reading, visual search, and scene perception: An overview. In: Rayner K, editor. *Cognitive and cultural*

- influences on eye movements. Tianjin, China: Tianjin People's Publishing House; 2008. pp. 3–33.
86. Privitera CM, Carney T, Klein S, Aguilar M. Analysis of microsaccades and pupil dilation reveals a common decisional origin during visual search. *Vision Res.* 2014;95: 43–50. doi:10.1016/j.visres.2013.12.001
 87. Rucci M, Iovin R, Poletti M, Santini F. Miniature eye movements enhance fine spatial detail. *Nature.* 2007;447: 852–855. doi:10.1038/nature05866
 88. Ko H, Poletti M, Rucci M. Microsaccades precisely relocate gaze in a high visual acuity task. *Nat Neurosci.* 2010;13: 1549–1553. doi:10.1038/nn.2663
 89. Willeke KF, Tian X, Buonocore A, Bellet J, Ramirez-Cardenas A, Hafed ZM. Memory-guided microsaccades. *Nat Commun.* 2019;10: 3710. doi:10.1038/s41467-019-11711-x
 90. Hafed ZM, Clark JJ. Microsaccades as an overt measure of covert attention shifts. *Vision Res.* 2002;42: 2533–2545. doi:10.1016/S0042-6989(02)00263-8
 91. Engbert R, Kliegl R. Microsaccades uncover the orientation of covert attention. *Vision Res.* 2003;43: 1035–1045. doi:10.1016/S0042-6989(03)00084-1
 92. Siegenthaler E, Costela FM, McCamy MB, Di Stasi LL, Otero-Millan J, Sonderegger A, et al. Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience.* 2014;39: 287–294. doi:10.1111/ejn.12395
 93. Pastukhov A, Braun J. Rare but precious: Microsaccades are highly informative about attentional allocation. *Vision Res.* 2010;50: 1173–1184. doi:10.1016/j.visres.2010.04.007
 94. Wen W, Yamashita A, Asama H. Divided Attention and Processes Underlying Sense of Agency. *Front Psychol.* 2016;7. doi:10.3389/fpsyg.2016.00035

95. Taylor JG. Does the corollary discharge of attention exist? *Conscious Cogn.* 2012;21: 325–339. doi:10.1016/j.concog.2011.09.018
96. García B, Gortázar F, Gallego M, Hines A. Assessment of QoE for Video and Audio in WebRTC Applications Using Full-Reference Models. *Electronics (Basel)*. 2020;9. doi:10.3390/electronics9030462
97. Yokokohji Y, Imaida T, Yoshikawa T. Bilateral teleoperation under time-varying communication delay. *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat No99CH36289)*. 1999. pp. 1854–1859 vol.3. doi:10.1109/IROS.1999.811748
98. Kim E, Peysakhovich V, Roy RN. Impact of communication delay and temporal sensitivity on perceived workload and teleoperation performance. *ACM Symposium on Applied Perception 2021*. New York, NY, USA: Association for Computing Machinery; 2021. doi:10.1145/3474451.3476233
99. Sasaki Y. Evaluation of Operability in Tele-operation Control of a Hydraulic Cylinder by the Packet Compensation Method. *Journal of MMIJ*. 2022;138: 149–159. doi:10.2473/journalofmmij.138.149
100. Musicant O, Botzer A, Shoval S. Effects of simulated time delay on teleoperators' performance in inter-urban conditions. *Transp Res Part F Traffic Psychol Behav*. 2023;92: 220–237. doi:10.1016/j.trf.2022.11.007
101. Scholcover F, Gillan DJ. Temporal Sensitivity and Latency During Teleoperation: Using Track Clearance to Understand Errors in Future Projection. *Hum Factors*. 2023;65: 260–274. doi:10.1177/00187208211011842
102. Timman S, Landgraf M, Haskamp C, Lizy-Destrez S, Dehais F. Effect of time-delay on lunar sampling tele-operations: Evidences from cardiac, ocular and

- behavioral measures. *Appl Ergon.* 2023;107: 103910.
doi:10.1016/j.apergo.2022.103910
103. Kim J, Yoshida T. Sense of agency at a gaze-contingent display with jittery temporal delay. *Front Psychol.* 2024;15. doi:10.3389/fpsyg.2024.1364076
 104. Sato A. Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition.* 2008;110: 74–83. doi:10.1016/j.cognition.2008.10.011
 105. Shimada S, Qi Y, Hiraki K. Detection of visual feedback delay in active and passive self-body movements. *Exp Brain Res.* 2010;201: 359–364.
doi:10.1007/s00221-009-2028-6
 106. Hon N. Attention and the sense of agency: A review and some thoughts on the matter. *Conscious Cogn.* 2017;56: 30–36. doi:10.1016/j.concog.2017.10.004
 107. Normoyle A, Guerrero G, Jörg S. Player perception of delays and jitter in character responsiveness. *Proceedings of the ACM Symposium on Applied Perception.* New York, NY, USA: Association for Computing Machinery; 2014. pp. 117–124. doi:10.1145/2628257.2628263
 108. Bolot J-C. End-to-end packet delay and loss behavior in the internet. *SIGCOMM Comput Commun Rev.* 1993;23: 289–298.
doi:10.1145/167954.166265
 109. Jiang W, Schulzrinne H. Modeling of packet loss and delay and their effect on real-time multimedia service quality. *Proc NOSSDAV.* 2000. pp. 1–10.
 110. Farrer C, Franck N, Georgieff N, Frith CD, Decety J, Jeannerod M. Modulating the experience of agency: a positron emission tomography study. *Neuroimage.* 2003;18: 324–333. doi:10.1016/S1053-8119(02)00041-1

111. Sirigu A, Daprati E, Ciancia S, Giraux P, Nighoghossian N, Posada A, et al. Altered awareness of voluntary action after damage to the parietal cortex. *Nat Neurosci.* 2004;7: 80–84. doi:10.1038/nn1160
112. Dale AM. Optimal experimental design for event-related fMRI. *Hum Brain Mapp.* 1999;8: 109–114. doi:10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W
113. Dale AM, Buckner RL. Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp.* 1997;5: 329–340. doi:10.1002/(SICI)1097-0193(1997)5:5<329::AID-HBM1>3.0.CO;2-5
114. Nonyane BAS, Theobald CM. Design sequences for sensory studies: Achieving balance for carry-over and position effects. *British Journal of Mathematical and Statistical Psychology.* 2007;60: 339–349. doi:https://doi.org/10.1348/000711006X114568
115. Fonov VS, Evans AC, McKinstry RC, Almli CR, Collins DL. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage.* 2009;47: S102. doi:10.1016/S1053-8119(09)70884-5
116. Worsley KJ, Friston KJ. Analysis of fMRI Time-Series Revisited—Again. *Neuroimage.* 1995;2: 173–181. doi:10.1006/nimg.1995.1023
117. Friston KJ, Glaser DE, Henson RNA, Kiebel S, Phillips C, Ashburner J. Classical and Bayesian Inference in Neuroimaging: Applications. *Neuroimage.* 2002;16: 484–512. doi:10.1006/nimg.2002.1091
118. Friston KJ, Williams S, Howard R, Frackowiak RSJ, Turner R. Movement-Related effects in fMRI time-series. *Magn Reson Med.* 1996;35: 346–355. doi:10.1002/mrm.1910350312

119. Neuromorphometrics, Inc. [cited 9 Jul 2024]. Available:
<https://www.neuromorphometrics.com/>
120. Lancaster JL, Rainey LH, Summerlin JL, Freitas CS, Fox PT, Evans AC, et al. Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Hum Brain Mapp.* 1997;5: 238–242. doi:10.1002/(SICI)1097-0193(1997)5:4<238::AID-HBM6>3.0.CO;2-4
121. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, et al. Automated Talairach Atlas labels for functional brain mapping. *Hum Brain Mapp.* 2000;10: 120–131. doi:10.1002/1097-0193(200007)10:3<120::AID-HBM30>3.0.CO;2-8
122. Esterman M, Tamber-Rosenau BJ, Chiu Y-C, Yantis S. Avoiding non-independence in fMRI data analysis: Leave one subject out. *Neuroimage.* 2010;50: 572–576. doi:10.1016/j.neuroimage.2009.10.092
123. Brett M, Anton J-L, Valabregue R, Poline J-B. Region of interest analysis using an SPM toolbox. 8th international conference on functional mapping of the human brain. Sendai, Japan; 2002. p. 497. Available: <https://marsbar-toolbox.github.io/>
124. Wen W, Imamizu H. The sense of agency in perception, behaviour and human–machine interactions. *Nature Reviews Psychology.* 2022;1: 211–222. doi:10.1038/s44159-022-00030-6
125. Ohata R, Asai T, Kadota H, Shigemasu H, Ogawa K, Imamizu H. Sense of Agency Beyond Sensorimotor Process: Decoding Self-Other Action Attribution in the Human Brain. *Cerebral Cortex.* 2020;30: 4076–4091. doi:10.1093/cercor/bhaa028

126. Xu J, Wang J, Fan L, Li H, Zhang W, Hu Q, et al. Tractography-based Parcellation of the Human Middle Temporal Gyrus. *Sci Rep*. 2015;5: 18883. doi:10.1038/srep18883
127. Laube I, Kamphuis S, Dicke PW, Thier P. Cortical processing of head- and eye-gaze cues guiding joint social attention. *Neuroimage*. 2011;54: 1643–1653. doi:10.1016/j.neuroimage.2010.08.074
128. Shimada S. Multisensory and Sensorimotor Integration in the Embodied Self: Relationship between Self-Body Recognition and the Mirror Neuron System. *Sensors*. 2022;22. doi:10.3390/s22135059
129. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 2002;3: 201–215. doi:10.1038/nnr755
130. Fox M, Corbetta M, Snyder A, Vincent J, Raichle M. Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc Natl Acad Sci USA* 103: 10046-10051. *Proc Natl Acad Sci U S A*. 2006;103: 10046–10051. doi:10.1073/pnas.0604187103
131. Wai AAP, Goh Z, Foo SD, Guan C. A Study of SSVEP Responses in Case of Overt and Covert Visual Attention with Different View Angles. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2018. pp. 3847–3852. doi:10.1109/SMC.2018.00652
132. de Lissa P, Caldara R, Nicholls V, Mielliet S. In pursuit of visual attention: SSVEP frequency-tagging moving targets. *PLoS One*. 2020;15: e0236967-. doi:10.1371/journal.pone.0236967

133. Haggard P, Clark S. Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition*. Academic Press Inc.; 2003. pp. 695–707. doi:10.1016/S1053-8100(03)00052-7
134. Humphreys GR, Buchner MJ. Magnitude estimation reveals temporal binding at super-second intervals. *J Exp Psychol Hum Percept Perform*. 2009;35: 1542–1549. doi:10.1037/a0014492

Appendix 1.2.

The task error rates were calculated for each participant, target, and condition. Error rates tended to be highest in the playback condition (Table 1.4). In this playback condition, participants frequently failed to inspect all eight surrounding characters because the recorded gaze behavior often did not encompass the entire character array.

Table 1.4. The task error rate for each participant, target existence, and each condition.

Participant	Delayed window				Playback condition			No window condition		
	condition									
	Trial	Error			Trial	Error		Trial	Error	
	counts	trial	Error rate		counts	trial	Error rate	counts	trial	Error rate
		counts				counts			counts	
1	no-target	461	7	1.5%	29	1	3.4%	24	1	4.2%
	target	432	17	3.9%	31	6	19.4%	12	2	16.7%
	Total	893	24	2.7%	60	7	11.7%	36	3	8.3%
2	no-target	457	1	0.2%	34	0	0.0%	24	0	0.0%
	target	442	2	0.5%	26	1	3.8%	12	1	8.3%
	Total	899	3	0.3%	60	1	1.7%	36	1	2.8%
3	no-target	457	3	0.7%	35	0	0.0%	24	0	0.0%
	target	443	15	3.4%	25	1	4.0%	12	0	0.0%
	Total	900	18	2.0%	60	1	1.7%	36	0	0.0%
4	no-target	427	0	0.0%	33	0	0.0%	18	1	5.6%
	target	450	9	2.0%	27	1	3.7%	18	0	0.0%
	Total	877	9	1.0%	60	1	1.7%	36	1	2.8%
5	no-target	459	0	0.0%	32	0	0.0%	24	0	0.0%
	target	440	3	0.7%	28	9	32.1%	12	0	0.0%
	Total	899	3	0.3%	60	9	15.0%	36	0	0.0%
6	no-target	453	1	0.2%	32	0	0.0%	24	0	0.0%
	target	447	10	2.2%	28	5	17.9%	12	0	0.0%
	Total	900	11	1.2%	60	5	8.3%	36	0	0.0%
7	no-target	453	4	0.9%	32	3	9.4%	24	0	0.0%
	target	437	32	7.3%	28	15	53.6%	12	0	0.0%

	Total	890	36	4.0%	60	18	30.0%	36	0	0.0%
8	no-target	456	1	0.2%	36	0	0.0%	22	0	0.0%
	target	444	3	0.7%	24	6	25.0%	14	0	0.0%
	Total	900	4	0.4%	60	6	10.0%	36	0	0.0%

Appendix 1.3.

It was anticipated that combinations of English letters could similarly prevent attentional bias and maintain similar spatial frequencies. To determine the most suitable language for the experiment, a pilot study was conducted using pools of English letters (Table 1.5) and Chinese characters (Appendix 1.1). The purpose of the pilot test was to evaluate which set of stimuli—Chinese characters or English letters—would be more appropriate for the visual search tasks, which were performed with a gaze-contingent window under two specific conditions: (1) stimuli were either Chinese characters (Figure 1.1) or English letters (Figure 1.9), and (2) the window was presented with a temporal delay relative to the participant’s gaze position, with delays ranging from 0 ms to 500 ms in 100 ms increments. Each participant completed 30 trials per condition.

The pilot study showed that the average response time for tasks using English letters was about 2 seconds shorter than for those using Chinese characters (Figure 1.10). Given that the primary aim of this pilot study was to identify a stimulus set that would enable participants to control the gaze-contingent window for a sufficient duration to gather eye movement data, these findings indicated that Chinese characters were more appropriate for the experimental design. As a result, 239 Chinese characters were chosen to create the stimulus images used in the main experiment, as detailed in Appendix 1.1.

Table 1.5. List of letter pairs selected for the stimuli in the pilot experiments using the Courier New font. Four letter pairs were chosen for their visual similarity, and one pair was randomly selected from these. Stimuli and mask images were then generated using these pairs. In each stimuli image, one character served as the target, while eight different characters made up the uniform background (e.g., “b” as the target and “p” as the background). Mask images featured nine identical characters (e.g., all “b” or all “p”). For each letter pair, two stimuli images and two mask images were created. These were then

combined in a 2×2 configuration, resulting in four combinations per pair, yielding a total of 16 combinations.

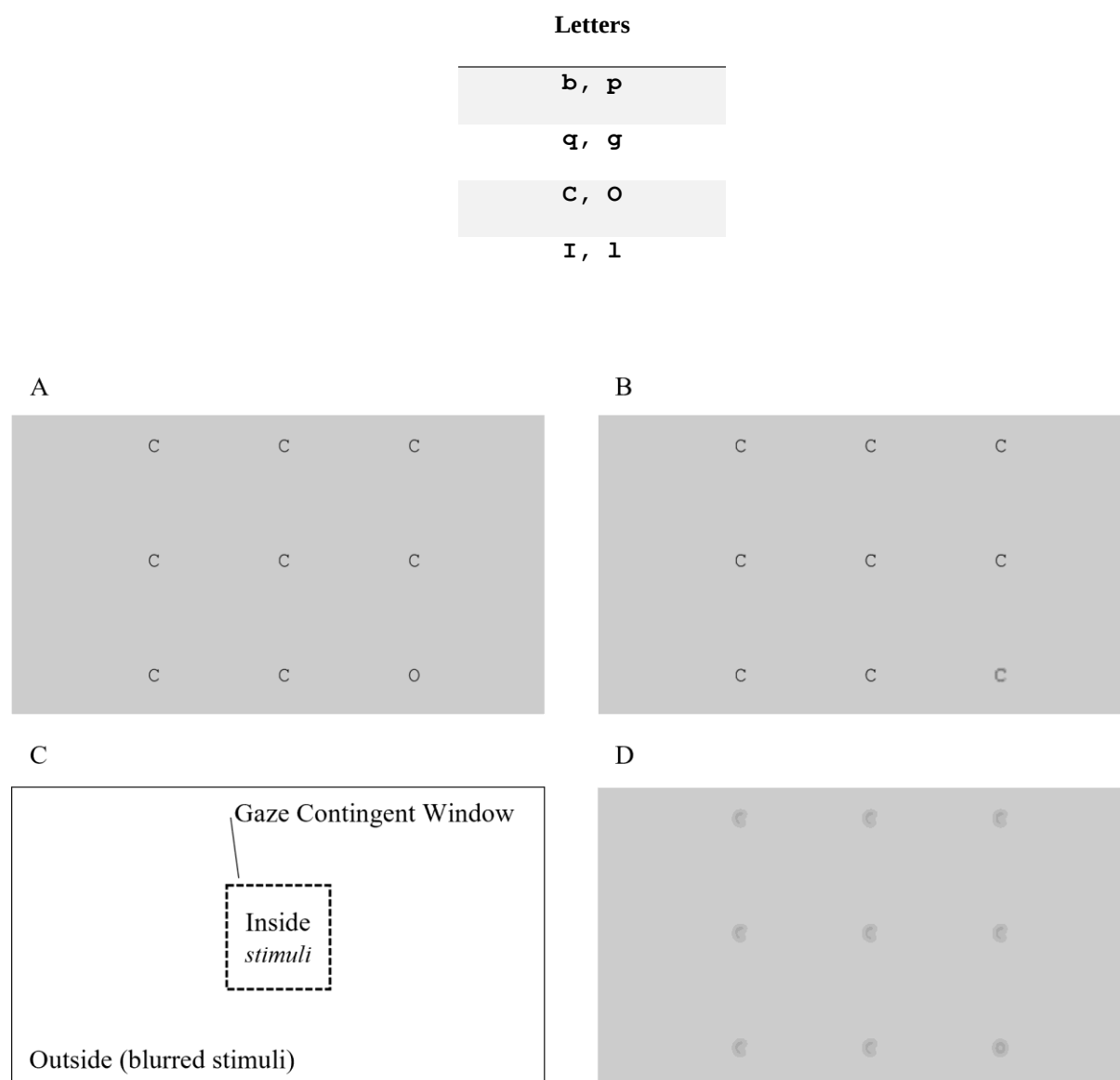


Figure 1.9. A) Stimuli image with the target character "O" located in the bottom right corner. B) Background image used for the experiment. C) A virtual square window was centered on the participant's fixation point, with the stimuli displayed inside the window and mask images shown outside. D) An example of the blurred stimuli, using a 10-pixel standard deviation Gaussian blur. Due to the significant reduction in visibility from the blur, the decision was made to use the mask image (B) rather than the blurred stimuli (D). Reproduced from the supplementary materials of Kim et al. [84].

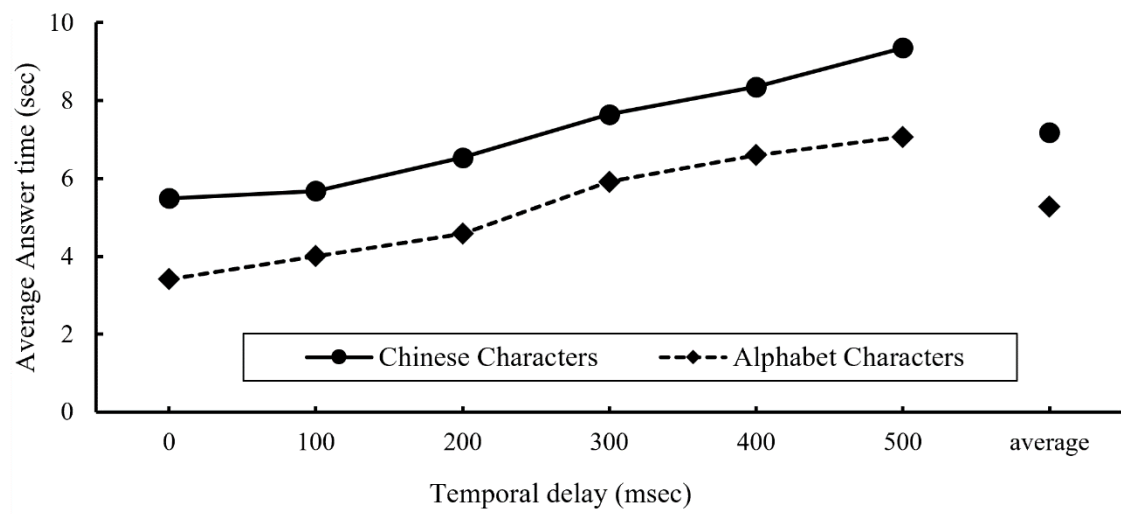


Figure 1.10. Average response time for English alphabet and Chinese characters visual search tasks. Reproduced from the supplementary materials of Kim et al. [84].

Appendix 1.4.

Whether participants might have strategically positioned their gaze between two adjacent characters to inspect both simultaneously was explored. Such a strategy could potentially reduce response times and introduce variability into the current analysis. To investigate this, the fixation count per trial for each defined area of interest on the stimuli was calculated (Table 1.6). The stimuli were divided into three key areas: 1) the gap region, where the gaze-contingent window could partially display parts of two Chinese characters when participants fixated on it (Figure 1.11, highlighted in black); 2) the word region, where only a single Chinese character would be fully visible within the window (Figure 1.11, highlighted in orange); and 3) the null region, where no character would be visible within the window (Figure 1.11, highlighted in blue). The average fixation count per participant (based on 1,080 trials per participant) was first calculated, and then the overall mean and standard deviation were computed, as shown in Table 1.6. The analysis revealed that very few fixations occurred in the gap region, indicating that participants did not position their gaze between two characters to try to view both at once. Moreover, fixations in the null region were minimal, suggesting that participants did not attempt to assess the temporal discrepancy by moving their gaze to the null region, even under high visibility conditions.

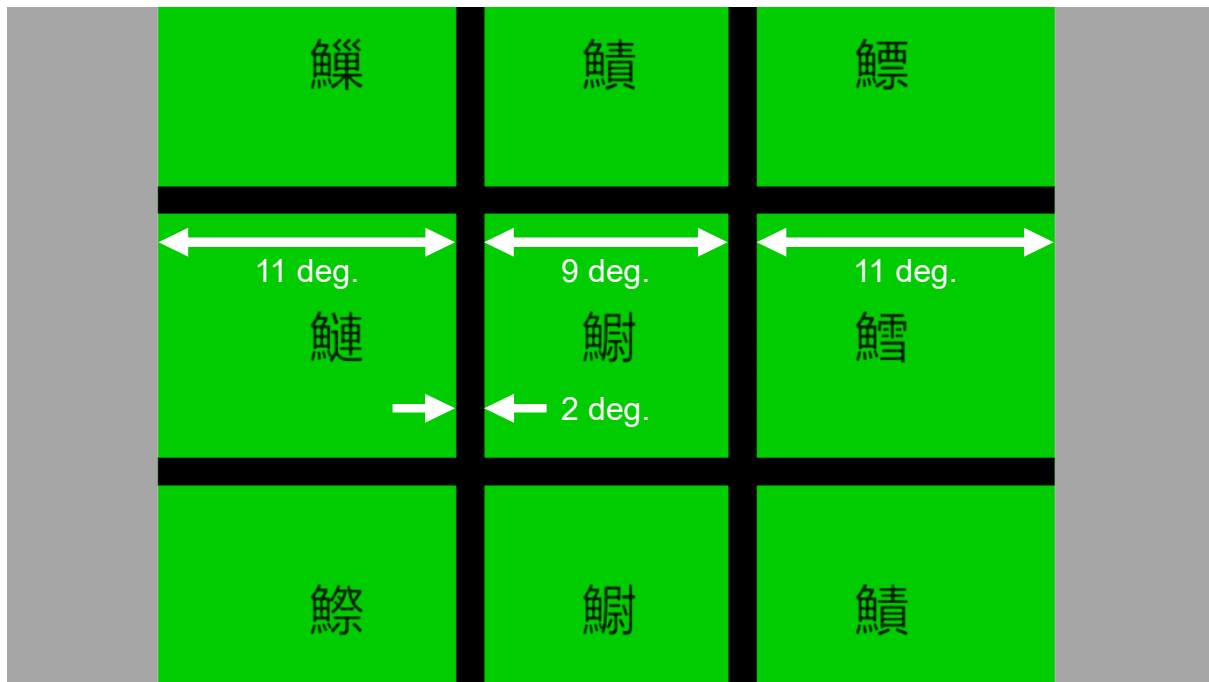
neighboring characters to view both simultaneously, potentially reducing response times and affecting eye movements.

0.1°, 30.0°/s, and 8000.0°/s² were set to identify saccades. Any fixations occurring outside the stimulus area, began immediately before or after a blink or lasted less than 120 ms were excluded from the current analyses.

I divided the stimuli into three areas of interest: 1) the gap region, where partial sections of two Chinese characters appeared within the gaze-contingent window when participants fixated (Appendix Figure 2.1, black); 2) the word region, where only one character was visible within the window (Appendix Figure 2.1, green); and 3) the null region, where no character was visible within the window (Appendix Figure 2.1, gray). The mean fixation count per trial in each area (Appendix Table 2.1) was calculated for each participant (270 trials). Very few fixations were observed in the gap region. This suggested that the participants did not intentionally gaze at two adjacent characters.

Appendix Table 2.1. The count of fixations per trial for each region. The means and standard deviations among the participants were calculated. The data showed few fixations on regions apart from the word region, suggesting that behaviors unrelated to the task barely occurred.

Regions	Mean	Standard deviation
Gap region	0.12	0.06
Word region	14.89	4.55
Null region	0.00	0.00



Appendix Figure 2.1. The stimuli were split into three distinct interest areas: 1) gap region, which reveals portions of two Chinese characters within the gaze-contingent window when the gaze was located (marked in black); 2) word region, rendering just one character within the window (highlighted in green); and 3) null region, which was unable to render any character within the window (denoted in grey). Reproduced from the supplementary materials of Kim et al. [103].

Appendix 2.2

The error rate per participant, condition, and the presence of the target were documented. The playback condition frequently exhibits the highest error rate. Participants struggled to inspect all eight surrounding characters in the playback condition because prerecorded gaze behavior often failed to reveal the entire character array.

Appendix Table 2.2. The task error rate for each participant, each condition, and target existence.

Participant	Stimuli	Delayed window condition			Playback condition			No window condition		
		Trial counts	Error trial counts	Error rate	Trial counts	Error trial counts	Error rate	Trial counts	Error trial counts	Error rate
1	no-target	120	2	1.67%	14	0	0.00%			
	target	117	3	2.56%	16	6	37.50%			
	Total	237	5	2.11%	30	6	20.00%			
2	no-target	120	0	0.00%	16	0	0.00%	15	0	0.00%
	target	119	4	3.36%	14	1	7.14%	15	0	0.00%
	Total	239	4	1.67%	30	1	3.33%	30	0	0.00%
3	no-target	126	0	0.00%	12	0	0.00%	15	0	0.00%
	target	114	3	2.63%	18	6	33.33%	15	0	0.00%
	Total	240	3	1.25%	30	6	20.00%	30	0	0.00%
4	no-target	120	0	0.00%	17	0	0.00%	15	0	0.00%
	target	120	0	0.00%	13	2	15.38%	15	0	0.00%
	Total	240	0	0.00%	30	2	6.67%	30	0	0.00%
5	no-target	123	1	0.81%	14	0	0.00%	15	0	0.00%
	target	117	0	0.00%	16	1	6.25%	15	0	0.00%
	Total	240	1	0.42%	30	1	3.33%	30	0	0.00%
6	no-target	119	0	0.00%	18	0	0.00%	15	0	0.00%
	target	121	5	4.13%	12	4	33.33%	15	0	0.00%
	Total	240	5	2.08%	30	4	13.33%	30	0	0.00%
7	no-target	124	2	1.61%	13	0	0.00%	15	0	0.00%
	target	116	2	1.72%	17	8	47.06%	15	0	0.00%
	Total	240	4	1.67%	30	8	26.67%	30	0	0.00%
8	no-target	119	0	0.00%	18	0	0.00%	16	0	0.00%
	target	121	0	0.00%	12	2	16.67%	14	0	0.00%
	Total	240	0	0.00%	30	2	6.67%	30	0	0.00%
9	no-target	121	0	0.00%	16	0	0.00%	15	0	0.00%
	target	119	3	2.52%	14	2	14.29%	15	0	0.00%

							%			
	Total	240	3	1.25%	30	2	6.67%	30	0	0.00%
10	no-target	122	0	0.00%	15	0	0.00%			
	target	118	3	2.54%	15	3	20.00%			
							%			
	Total	240	3	1.25%	30	3	10.00%			
							%			
11	no-target	125	1	0.80%	12	0	0.00%	15	0	0.00%
	target	115	2	1.74%	18	5	27.78%	15	0	0.00%
							%			
	Total	240	3	1.25%	30	5	16.67%	30	0	0.00%
							%			

Appendix 2.3

Appendix Table 2.3. Covariance parameters and fixed effects estimates for the authorship questionnaire, comparing the self to the delay response.

Covariance parameter	Estimate	Standard error	Z score	P value
Participants	1.347	0.627	2.147	.032
Fixed effects	Coefficient	Standard Error	T score	P value
Intercept	1.827	0.496	3.686	.001
Delay	-0.018	0.002	-8.913	.000

Appendix Table 2.4. Covariance parameters and fixed effects estimates for the authorship questionnaire, comparing the other to the delay response.

Covariance parameter	Estimate	Standard error	Z score	P value
Participants	0.708	0.474	1.494	.135
Fixed effects	Coefficient	Standard Error	T score	P value
Intercept	-3.387	0.507	-6.683	.000
Delay	0.000	0.002	0.122	.903

*Appendix Table 2.5. The results of the one-way multivariate ANOVA for the fitted values of slope (**a**) and threshold delay (**b**) within the delayed trials with the experiment type as a between-subjects factor.*

	Pillai's Trace Value	df	F	Error df	P	η_p^2	Noncent. Parameter	Observed Power
Intercept	.912	4.000	33.736	13.000	.000	.912	134.944	1.000
Experiment	.124	4.000	.459	13.000	.764	.124	1.838	.125

Appendix Table 2.6. The results of the one-way ANOVA with repeated measures on mean authorship rating score.

	df	F	Error df	P	η_p^2	Noncent. Parameter	Observed Power
Delay	1.303	65.974	13.025	.000	.868	85.934	1.000

Appendix Table 2.7. The results of the target $\times$ delay repeated-measures ANOVA on the response time

	df	F	Error df	P	η_p^2	Noncent. Parameter	Observed Power
Delay	1.734	33.814	17.341	.000	.772	58.635	1.000
Target	1.000	23.471	10.000	.001	.701	23.471	.991
Delay $\times$ Target	2.393	2.358	23.931	.108	.191	5.643	.468

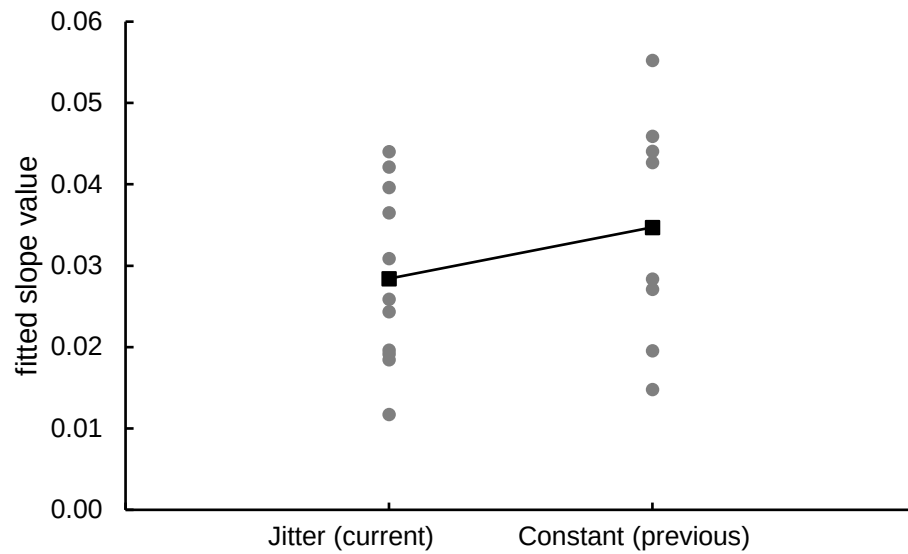
Appendix Table 2.8. The results of the one-way ANOVA with repeated measures on fixation counts per trial.

	df	F	Error df	P	η_p^2	Noncent. Parameter	Observed Power
Delay	1.751	.774	17.512	.460	.072	1.356	.154

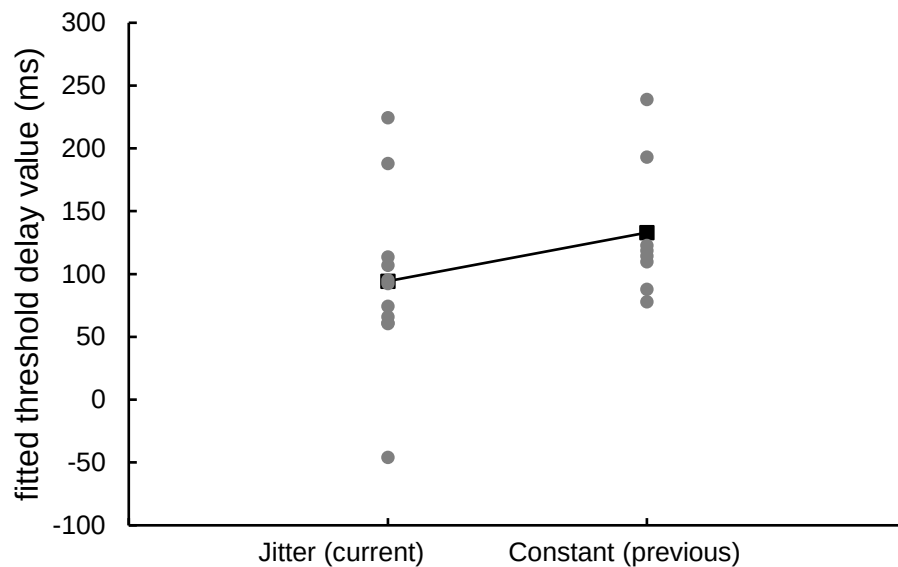
Appendix Table 2.9. The results of the two-way mixed ANOVA for the fixation count per second within the delayed trials. The delay variable was treated as a within-subjects factor, while the type of experiment (constant versus jittery) was considered a between-subjects factor.

Between-subject effect	df	F	P	η_p^2	Noncent. Parameter	Observed Power
Intercept	1.000	1245.617	.000	.987	1245.617	1.000
Experiment	1.000	.961	.341	.053	.961	.152
Error (Experiment)	17.000					
Within-subject effect	df	F	P	η_p^2	Noncent. Parameter	Observed Power
Delay	1.473	90.652	.000	.842	133.493	1.000
Delay $\times$ Experiment	1.473	.416	.603	.024	.613	.103
Error (Delay)	25.034					

Appendix 2.4



Appendix Figure 2.2. The grey circle dots represent raw data of the fitted slope values (a) of ‘self’ responses from both the current experiment ($N = 11$) and the previous experiment ($N = 8$). The black squares indicate the average value at each experiment. Reproduced from the supplementary materials of Kim et al. [103].



Appendix Figure 2.3. The grey circle dots represent raw data of the fitted threshold delay values (b) of ‘self’ responses from both the current experiment ($N = 11$) and the previous experiment ($N = 8$). The black squares indicate the average value at each experiment.

Reproduced from the supplementary materials of Kim et al. [103].

Achievements

Publication

- Kim, J., & Yoshida, T. (2024). Sense of agency at a temporally-delayed gaze-contingent display. *PLOS ONE*, doi: 10.1371/journal.pone.0309998.
- Kim, J., & Yoshida, T. (2024). Sense of agency at a gaze-contingent display with jittery temporal delay. *Frontiers in Psychology*, 15, 1364076.

International Conferences

- Kim, J., & Yoshida, T. (2024, August). Sense of agency at a gaze-contingent display with jittery temporal delay. Poster session prepared for the 46th European Conference on Visual Perception, Aberdeen, Scotland. (Presented by a colleague due to illness)
- Yasunaga, M., Kim, J., & Yoshida, T. (2024, August). Effects of temporal delay on task performance and sense of agency in continuous tracking task. Poster session for the 46th European Conference on Visual Perception, Aberdeen, Scotland. (Presented by a colleague)
- Kim, J., & Yoshida, T. (2023, November). Neural correlates of sense of agency in temporally delayed gaze-contingent display: an fMRI Study. Poster session presented at the Society for Neuroscience 2023, Washington D.C., USA.
- Kim, J., & Yoshida, T. (2022, August). Examining sense of agency in the jittery temporal interval between gaze shift and visual feedback using gaze-contingent multiresolution display. Poster session presented at the 44th European Conference on Visual Perception, Nijmegen, Netherlands.
- Kim, J., & Yoshida, T. (2022, August). The temporal order judgment between saccade and visual stimulation just after saccade. Poster session presented at the 21st European Conference on Eye Movements, Leicester, United Kingdom.

- Kim, J., & Yoshida, T. (2021, August). The critical temporal delay between gaze shift and visual feedback for maintaining a sense of agency – Gaze-contingent multiresolutional display study. Talk session presented at the 43rd European Conference on Visual Perception, online.

Domestic Conference

- Kim, J., & Yoshida, T. (2021, September). 視線移動に応じた多重解像度ディスプレイで制御感覚が失われる臨界描画遅延時間 (The critical temporal delay between gaze shift and visual feedback for maintaining a sense of agency in a gaze-contingent multiresolutional display). Talk session presented at the Vision Society of Japan 2021, online.