HOKKAIDO UNIVERSITY

| | |
|---|---|
| Title | Visual Factors Shaping Text Legibility and Preference in Virtual Reality : A Comparative Study of Native and Non-Native Speakers |
| Author(s) | 張, 慧丹 |
| Degree Grantor | 北海道大学 |
| Degree Name | 博士(情報科学) |
| Dissertation Number | 甲第16984号 |
| Issue Date | 2026-03-25 |
| DOI | https://doi.org/10.14943/doctoral.k16984 |
| Doc URL | https://hdl.handle.net/2115/99828 |
| Type | doctoral thesis |
| File Information | Huidan_Zhang.pdf, 全文 |

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

博士論文

# バーチャルリアリティー環境におけるテキスト判別性と嗜好性に影響する視覚要因の調査:

母語話者と非母語話者の比較研究

北海道大学 大学院情報科学院
情報科学専攻 情報理工学コース
ヒューマンコンピュータインタラクション研究室

張慧丹

**2026 年 2 月**

**Doctoral Thesis**

# Visual Factors Shaping Text Legibility and Preference in Virtual Reality:

***A Comparative Study of Native and Non-Native Speakers***

by

Zhang Huidan

Human-Computer Interaction Laboratory,
Course of Computer Science and Information Technology,
Division of Computer Science and Information Technology,
Graduate School of Information Science and Technology,
Hokkaido University
February 2026

# Contents


**1 Introduction** **1**

1.1 Background . . . 2

1.1.1 Evolution of Legibility Research . . . 2

1.1.2 Cross-Language Differences in Reading Behavior (L1 vs L2) . . . 4

1.2 Objective . . . 6

1.3 Structure . . . 7

**2 Method** **11**

2.1 Apparatus . . . 11

2.2 VR System Overview . . . 11

2.3 Parameters . . . 12

2.3.1 Native and Non-native . . . 12

2.3.2 Distance . . . 14

2.3.3 Glyph . . . 14

2.3.4 Font: font style and font weight . . . 19

2.4 Method . . . 20

**3 Perceiving Multilingual Text in Virtual Reality: Glyph Complexity and Font Effects on Preferred Viewing Distance** **23**

3.1 Related Work . . . 23

3.1.1 Psychophysics of distance perception and comfort in VR . . . 23

3.1.2 Design factors that influence distance-dependent readability in VR 24

3.1.3 Population differences in perceived distance to text . . . 24

3.2 Experiment Design . . . 25

3.2.1 Independent variable . . . 26

3.2.2 Task . . . 28

3.2.3 Procedure . . . 29

3.2.4 Data collection and analysis . . . 30

3.3 Results . . . 31

3.3.1 Japanese experiment . . . 31

3.3.2 Chinese experiment . . . 32

3.3.3 English experiment . . . 33

3.3.4 Result Summary . . . 34

3.4 Discussion of Key Findings . . . 35

3.4.1 Effect of Native Language on Perceptual Distance . . . 36

3.4.2 Visual compensation . . . 36
3.4.3 Design Implications for Complex Glyphs . . . 38
3.4.4 Subtle Effects of Font Weight . . . 39
3.5 Chapter Summary . . . 41

**4 Understanding Glyph Characteristics on the Legibility for Native and Non-Native Speakers in Virtual Reality 43**
4.1 Related Work: Cross-Linguistic Cognitive Differences in Native vs. Non-Native Language Processing . . . 44
4.1.1 Cognitive and Behavioral Disparities in L1 vs. L2 Processing . . . 44
4.1.2 Glyph Perception in Logographic vs. Alphabetic Systems . . . 44
4.1.3 Synthesis and Research Positioning . . . 45
4.2 Experiment Design . . . 46
4.2.1 Virtual Environment . . . 46
4.2.2 Independent variable . . . 46
4.2.3 Task . . . 50
4.2.4 Procedure . . . 50
4.2.5 Data Collection and Analysis . . . 51
4.3 Result . . . 52
4.3.1 Error Rates . . . 53
4.3.2 Invisibility Rates . . . 59
4.3.3 Task Completion Time . . . 63
4.4 Discussion of Key Findings . . . 69
4.4.1 Differences in Visual Recognition between L1 and L2 Participants 69
4.4.2 Educational and Design Implications . . . 71
4.5 Chapter Summary . . . 73

**5 Study on font preferences of native and non-native speakers in a virtual reality environment 75**
5.1 Related work . . . 76
5.1.1 Digital Factors Affecting Digital Reading . . . 76
5.1.2 Individual Factors Affecting legibility . . . 76
5.1.3 Cognitive Font Preferences of Native and Non-Native Speakers . . . 77
5.1.4 Summary . . . 77
5.2 Experiment Design . . . 77
5.2.1 Virtual Environment . . . 78
5.2.2 Independent Variables . . . 78
5.2.3 Task . . . 81
5.2.4 Procedure . . . 81
5.2.5 Data Analysis and Processing . . . 82
5.3 Results . . . 84
5.3.1 Semi-structured Interview Analysis . . . 84

5.3.2 Analysis of Individual Factors . . . . . 86
5.4 Discussion of Key Findings . . . . . 97
5.4.1 Distinct font preferences for Native and Non-Native speakers across language groups . . . . . 97
5.4.2 Impact of Font Style and Weight on Preferences and Design Implications . . . . . 99
5.4.3 Summary of Findings . . . . . 100
5.5 Chapter Summary . . . . . 103

**6 Discussion 104**
6.1 Operationalization of VR text-legibility findings . . . . . 104
6.1.1 Language-Independent VR Text Legibility Guidelines . . . . . 104
6.1.2 Initial attempts at future work: VTLS- A real-time legibility system for VR text based on VR guidelines . . . . . 105
6.2 Perceptual–Cognitive Transitions between Native and Non-Native Readers in VR . . . . . 110
6.2.1 Visual Requirements . . . . . 110
6.2.2 Role of Experience . . . . . 110
6.2.3 Influence of Cultural Exposure . . . . . 111
6.2.4 Initial attempts at future work: Design Implications for Native and Non-Native Readers . . . . . 111
6.3 Limitations and Directions for Future Research . . . . . 113

**7 Conclusion 115**

**A List of Research Achievements 134**
A.1 Journal Articles (peer-reviewed) . . . . . 134
A.2 International Conference Papers (peer-reviewed) . . . . . 134

**B Power Analysis of GLM Models 135**

# Chapter 1 Introduction

Walking requires no conscious thought, breathing requires no deliberation – indeed, many everyday actions are driven by perception: diary studies estimate that about 40% of daily behaviors are habitual rather than deliberative [1]. Many fields have recognized and theorized this phenomenon of unthinking, intuitive behavior, each with its own terminology. This accords with dual-process accounts in psychology, where fast, automatic (Type 1) processes often guide perception and action with minimal effort [2].

In behavioral economics, the influential nudge theory (pioneered by Richard Thaler, who was awarded the 2017 Nobel Prize in Economic Sciences) is built on the idea that small changes in how choices are presented can unconsciously "nudge" people into better decisions [3]. In the design field, Japanese industrial designer Naoto Fukasawa introduced the concept of "design without thought", emphasizing that the best designs leverage users' unconscious behaviors, an approach epitomized by MUJI's minimalistic products [4]. In the HCI(Human Computer Interaction) community, a parallel principle is encapsulated by Steve Krug's famous maxim "Don't make me think", which argues that well-designed interfaces should allow users to accomplish tasks unconsciously, without pausing to mentally process each step [5–7]. From a neuroscientific perspective, these intuitive behaviors can be seen as the result of memory and learning processes deeply engrained in the brain. The hippocampus (a brain structure located deep in the temporal lobe) plays a crucial role in consolidating our experiences into long-term memory [8]. The accumulated knowledge and experiences in our long-term memory continuously shape our perceptions and guide our actions at a subconscious level [9].

When reaching for a spoon to eat ice cream, people usually pick the utensil that feels most comfortable rather than the flashiest. That seemingly trivial choice reflects many factors—age and motor skills, context (standing vs. seated), temperature and texture, lighting, and cultural norms—and it is resolved without deliberation. What feels "obvious" is thus the visible tip of an iceberg shaped by prior knowledge and sensory cues operating below awareness.

Following this principle, user-centered design begins with careful observation of users' natural behaviors, feelings, and perceptual routines. In practice, systems should align with ingrained habits and perceptual tendencies so that interaction requires minimal effort. Because perception is the gatekeeper of cognition, it is necessary to understand what makes information perceptually available in the first place: how basic visual factors jointly determine what is visible, clear, and discriminable.Accordingly, this dissertation considers text in virtual reality (VR) and concentrates on a small set of first-order

visual determinants. These include viewing geometry (distance/scale), the structure and complexity of the stimulus, and the form/contrast of its rendering. Clarifying how they jointly set visibility, clarity, and identifiability is necessary before drawing conclusions about comprehension or design.

## 1.1 Background

### 1.1.1 Evolution of Legibility Research

#### Print work to screen work

Research on text legibility has a rich history spanning traditional print, electronic screens, and now immersive displays. In the print era, systematic legibility studies date back over a century. Pioneering experimental work by Tinker and Paterson established core print-legibility guidelines [10–13], showing effects of type size, line length, interline spacing (leading), and typeface style on reading performance. These print-based investigations revealed, for example, that serif typefaces and certain x-height proportions improved reading speed in long form text, and they set baselines such as the familiar 10–12 point text size for body copy at typical reading distances.

With the advent of computer displays, researchers examined whether print-based legibility principles would carry over to low–pixel-density cathode-ray-tube (CRT) displays. Early evidence indicated that inadequate image quality directly impaired legibility and comfort, and that reading from screens was typically slower than from paper [14]. In particular, when mainstream systems operated at approximately 72 dpi, text often appeared jagged and fine typographic details (e.g., serifs, thin crossbars) could blur or disappear [15]. Consequently, 1990s screen-typography guidance favored larger body sizes and sans-serif faces to mitigate the poor rendering of fine strokes on coarse pixel grids [16].

As anti-aliasing and subpixel rendering [17] and the introduction of "retina-class" panels [18] pushed pixel densities into the 200–300 ppi range, previously critical differences —such as serif versus sans-serif at small sizes—tended to diminish [19]. Users could read smaller on-screen text comfortably, and type design moved beyond early "pixel-proof" faces such as Verdana [20].

#### XR work

Against this backdrop, immersive displays in eXtended Reality (XR), including Virtual Reality (VR) head-mounted displays, reopen the legibility question: While many fundamentals of readability still apply (text size, contrast, typeface choice, etc.), the VR medium brings unique optical and ergonomic factors that can significantly alter previously established perceptual thresholds.

First, spatial distance and viewing angle are perceived differently in head-mounted

VR. Common HMD rendering practices—such as world-locked text (anchoring text to the virtual world rather than to the screen)—introduce dynamic blur: as the user moves their head, the text shifts across the pixel grid from frame to frame, which effectively reduces sharpness unless special supersampling is applied [21, 22]. There is no direct analogue of this phenomenon on 2D displays or in book-like reading.

Second, most head-mounted displays (HMDs) provide a much narrower field of view (FoV) than natural vision. Manufacturer specifications commonly target on the order of 90–110° diagonally, but effective/usable FoV can be noticeably smaller in practice, and widening FoV generally trades off against angular resolution for a fixed panel resolution and optics. As a result, when text is spread across a large virtual workspace, users may need more head or eye rotations than on a large desktop monitor to keep lines within the high-acuity region [23, 24]. Design guidance therefore recommends keeping critical text comfortably within the central FoV [25].

Furthermore, clarity across the image is spatially non-uniform in typical HMDs. Wide-angle lenses introduce radial distortion and chromatic aberration; rendering pipelines must pre-warp (apply barrel distortion) per eye to cancel the lens' s pincushion distortion, but this resampling can reduce sharpness, and residual aberrations leave center–periphery differences in legibility—text pushed toward the edges can look stretched or blurred compared with the center [26]. Consistent with this, psychophysical tools tailored for XR show that flat-panel legibility thresholds do not transfer directly to VR and should be re-measured under VR viewing (e.g., world-fixed text, off-axis viewing) [27, 28].

**What other jobs are worth doing?**

To map what has—and has not—been studied, a rapid scoping sweep was conducted on Crossref and PubMed using the query family (legibility OR readability) AND (text OR font OR typeface OR glyph) AND (virtual reality OR VR OR head-mounted display OR HMD OR augmented reality OR AR), together with an additional print/typography query. Peer-reviewed studies on immersive head-mounted displays (VR/AR) and on print that experimentally evaluated text/glyph legibility or reading performance were retained. Each eligible record was coded along three axes: (A) script/language, (B) visual/text factors (e.g., size/viewing angle, typeface/weight, glyph complexity, contrast/polarity, spacing, placement/anchor/angle, motion/instability, device pixels-per-degree/resolution), and (C) population/context (general adults, older or low-vision readers, dynamic or extreme viewing situations, cross-language settings). For the factor categories, a recent XR readability taxonomy was adopted so that the coding scheme aligns with the existing literature [29].

At a glance (counts produced by the script):

(A) Script/language 1.1. The corpus remains heavily Latin/English-centric; non-alphabetic scripts (e.g., Chinese/Japanese) and Arabic appear but are comparatively sparse.

Table. 1.1: Script/language distribution of VR legibility papers (counts)

| Script/Language | Latin | Japanese | Korean | Chinese | Arabic | Other |
|---|---|---|---|---|---|---|
| **Paper count** | 14422 | 48 | 46 | 45 | 20 | 11 |

(B) Factors studied 1.2. Work clusters around text size / angular height / virtual distance; studies on typeface/weight/glyph are fewer but growing; contrast/polarity, spacing, placement/anchor/angle, motion/instability, and device PPD/resolution form smaller yet important strands.

Table. 1.2: Factors studied in VR legibility papers (counts)

| Factor | Device | Review | Motion | Glyph | Spacing | Contrast | Typeface | Angle | Distance |
|---|---|---|---|---|---|---|---|---|---|
| **Paper count** | 2020 | 896 | 502 | 324 | 246 | 191 | 188 | 108 | 93 |

(C) Populations/contexts 1.3. Participants are mostly general adults; older/low-vision, dynamic/extreme (movement, low luminance), and explicit L1–L2 comparisons are under-represented—precisely the gaps this dissertation targets.

Table. 1.3: Populations/contexts in VR legibility papers (counts)

| Population | General | Dynamic | Low-vision | Cross-language |
|---|---|---|---|---|
| **Paper count** | 13788 | 563 | 225 | 16 |

These observations motivate the present dissertation' s focus: script-sensitive perceptual measurement (preferred viewing distance, glyph recognition, and font preference) across L1 vs. L2 readers in VR—precisely where current evidence is thinnest. These gaps point directly to a measurement agenda for multilingual VR reading. Before turning to our VR experiments, it is essential to ground the problem in what psycholinguistics and vision science already know about L1–L2 differences on traditional (2D) displays.

### 1.1.2 Cross-Language Differences in Reading Behavior (L1 vs L2)

#### Differences in cognitive frameworks between L1 and L2 readers

Research on reading commonly distinguishes *top–down* and *bottom–up* processing. In brief, skilled first-language (L1) readers can rely on high-quality lexical representations and contextual predictions (top–down), which helps stabilize recognition even when the visual signal is somewhat degraded. By contrast, second-language (L2) readers—especially at intermediate proficiency—tend to depend more on bottom-up evidence from letters and word forms and make less efficient use of context; this contrast is often framed as a continuum of lexical quality [30, 31]. For example, a fluent L1 reader of English can skim quickly and leverage context, whereas an L2 reader of the same text may engage in more word-by-word decoding that focuses on surface form.

These processing differences are visible in eye movements: compared with L1 read-

ers, L2 readers typically show longer fixation durations, shorter saccades, lower skipping rates, and sometimes more regressions—patterns consistent with slower lexical access and a stronger reliance on sublexical detail [32, 33]. Such differences matter for interface design on conventional screens: a typeface or layout that feels optimal for native readers may not be optimal for non-native readers if the latter benefit from greater visual clarity or are more easily distracted by ornamental features. HCI studies reflect this perspective by exploring adaptive support for non-native reading (e.g., interfaces that add machine-translation snippets when eye movements signal difficulty) and by showing that font effectiveness varies substantially across individuals, suggesting that simpler, clearer presentations or individualized font choices can be helpful [34, 35].

### Differences across logographic (character-based), syllabic, and alphabetic scripts

Prior work indicates that readers of logographic/character-based scripts (e.g., Chinese characters and Japanese kanji ) develop fine-grained sensitivity to complex character shapes, component (radical) structure, and visual similarity among glyphs; by contrast, readers of alphabetic scripts are tuned to letter-sequence patterns and position coding that support rapid assembly of word identities [36–39]. In Chinese, for example, both the functional role and the position of radicals contribute to identification, and visual similarity is strongly constrained by shared structure/components [38, 40]. Psychophysical evidence further shows that the diagnostic visual information differs by script: the minimum spatial-frequency content needed for accurate recognition of Chinese characters is not the same as for alphabetic letters, reflecting differences in two-dimensional composition and complexity [41].

By contrast, syllabic scripts such as Japanese kana (hiragana/katakana) are highly regular mappings from symbols to mora/syllable-sized phonology. Empirical comparisons indicate higher orthographic–phonological consistency for kana than for kanji and differences in how phonological and semantic information are weighted across the two Japanese scripts [42]. These properties make kana reading more tolerant of letterlike noise than kanji in some tasks, while kanji often place greater demands on whole-word or component-structure processing [43, 44].

For alphabetic reading, models and data emphasize order-sensitive orthographic coding (e.g., sensitivity to relative letter positions and orthotactic patterns), which helps explain fluent recognition despite letter-level noise [39]. Sensitivity to diacritics, however, is language-specific: in Spanish, omitting accent marks often produces little or no cost in word identification, whereas in German diacritics (e.g., umlauts) are integral to the lexicon and their omission can hinder access [45]. These contrasts illustrate that what counts as visually diagnostic varies across writing systems and languages, a point with direct implications for script-agnostic interface guidelines.

Cross-language studies also suggest that even highly proficient bilinguals do not fully

converge on monolingual native patterns in all respects. L2 readers can be more sensitive to changes in surface linguistic form than L1 readers (e.g., when wording is altered but meaning is preserved), consistent with heavier weighting of form-level information under effortful decoding [46]. Taken together, these findings motivate treating script and language background as meaningful factors when designing and evaluating text presentation.

**What other jobs are worth doing?**

Despite steady progress on text readability for AR/VR, very little is known about how native (L1) vs. non-native (L2) differences manifest under immersive, near-eye presentation. Recent reviews synthesize tunable text parameters and device constraints, but they largely treat display factors rather than cross-language reader groups [29]. Meanwhile, the VR medium imposes its own optical limits—e.g., angular resolution and field-of-view trade-offs, lens distortion, and non-uniform pixels-per-degree across the visual field —that can shift the perceptual thresholds established on flat screens [23, 24].

These constraints intersect with well-documented processing differences between L1 and L2 reading: L1 readers typically benefit more from top-down predictability, whereas L2 readers rely more on bottom-up visual evidence. This raises open questions for immersive text: Do L2 readers, who already depend more on fine-grained visual detail, require relatively larger effective angular sizes in VR (for a fixed physical text size, a closer viewing distance produces a larger visual angle, whereas a farther distance reduces it)? Do blur, contrast loss, or geometric distortion disproportionately tax one group, or can certain viewing regimes level performance between L1 and L2? To date, XR legibility work has begun to standardize measurement methods (e.g., single-letter thresholds across headsets) but has seldom contrasted L1 and L2 populations directly [27]. Addressing this gap, the present dissertation examines script-sensitive perception in VR across language backgrounds under a deliberately constrained set of viewing conditions: seated observers in a stable, indoor VR setting, with high-contrast text on a uniform background and no additional motion or task demands beyond reading. By first establishing how L1 and L2 readers perceive and prefer text under these controlled conditions, the results provide a baseline for future work that will extend to more complex and ecologically varied VR scenarios.

## 1.2 Objective

Given the above background, design guidelines for text legibility that were established for print or flat screens may not transfer directly to immersive media—especially for diverse user populations. Parameters traditionally deemed “optimal” (e.g., font weight, serif vs. sans-serif, glyph complexity) need to be re-examined in Virtual Reality (VR), where near-eye optics and viewing geometry can change perceptual thresholds. In par-

ticular, subtle differences in perceptual strategy between native and non-native readers may translate into different requirements for a comfortable and effective VR reading experience. A one-size-fits-all approach to typography is therefore likely to be suboptimal; instead, a nuanced account of how language background interacts with visual text factors under VR conditions is needed.

The overarching objective of this dissertation is to investigate how native vs. non-native language readership influences text perception in VR across multiple dimensions of visual presentation. This objective is organized around three core issues:

1. **Perceived Text Size / Viewing Distance.** How does language familiarity affect comfortable reading distance or required text size in VR? Prior work outside VR has not established whether L2 readers need proportionally larger text (or closer viewing) to achieve equal clarity. This dissertation will compare native and non-native readers' preferred viewing distances for text panels in VR and examine how font style and weight modulate these preferences.

2. **Glyph Recognition and Font–Glyph Complexity.** Do L1–L2 differences in visual reading strategy extend to character- or word-level recognition in VR, particularly as font style and glyph complexity vary? This dissertation will test how glyph complexity (e.g., simple vs. intricate shapes across scripts or typefaces) and glyph similarity (confusable letters/characters) impact recognition at the limits of legibility (e.g., small angular size or greater distance), and whether non-native readers are differentially affected.

3. **Font Aesthetics and User Preference.** How do native and non-native readers differ in subjective font preferences in VR when reading and comprehension demands are minimized? To avoid conflating aesthetics with difficulty, this dissertation will collect comparative preferences under matched content and viewing conditions. This allows assessment of whether language background shifts what is considered "comfortable" or "pleasing" to read in VR.

By systematically studying these three aspects through controlled experiments, the dissertation aims to build a theoretical and empirical basis for adaptive, language-sensitive typography in immersive environments. The results are intended to inform VR systems that can dynamically adjust text rendering—such as font selection, sizing, or styling—to better suit a reader's linguistic background, thereby improving legibility and reducing cognitive load for multilingual user groups in education, communication, and everyday VR use.

## 1.3 Structure

Figure 1.1 provides a schematic overview of the seven chapters. The dissertation is organized around a controlled, comparative framework: all experiments were conducted on

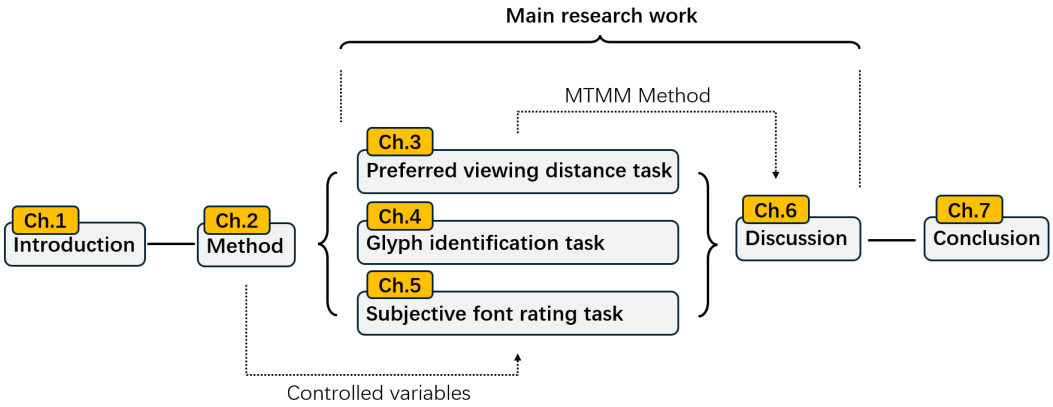



Fig. 1.1: Overview of the dissertation structure.

the same VR device, under the same laboratory conditions and calibrated distance/size settings, and then combined within a Multi-Trait Multi-Method (MTMM) perspective. Two traits are examined—legibility and preference—using three complementary methods: a preferred viewing distance task, a glyph identification task, and a subjective font rating task. Across these methods, native Chinese, Japanese, and English readers perform tasks in both their first language (L1) and a less familiar second language (L2), allowing visual factors and language background to be disentangled from hardware or contextual differences.

**Chapter 1: Introduction**

Chapter 1 positions the thesis at the intersection of VR legibility, cross-script reading, and L1–L2 differences. It traces the shift from print and 2D displays to near-eye VR displays, emphasizing how headset optics (field of view, distortion, pixels-per-degree) alter the physical conditions under which "clear enough to read" must be defined. The chapter reviews psycholinguistic and eye-movement findings showing that legibility thresholds depend on both the reader (L1 vs. L2) and the script (alphabetic vs. character-based), and identifies the lack of systematic, cross-script VR studies as a central gap. On this basis it motivates the comparative strategy adopted in the thesis, articulates the three core research issues (viewing distance, glyph recognition, and font preference), and introduces the overall MTMM-style structure and the idea of a legibility-aware VR text rendering system.

**Chapter 2: Methods**

Chapter 2 establishes the shared methodological framework used throughout the experimental chapters. It describes the VR apparatus and calibration procedures for distance and angular size, the participant recruitment and language-grouping criteria, and the construction of the multilingual text stimuli. The chapter details how fonts and scripts were selected and paired (e.g., serif/sans-serif, light/bold), how glyph complexity

and similarity were quantified, and how trial structure and logging were implemented in Unity. By defining these variables and procedures once and applying them consistently across Chapters 3–5, Chapter 2 provides a common reference frame that makes the subsequent experiments directly comparable.

**Chapter 3: Perceiving Multilingual Text in Virtual Reality: Glyph Complexity and Font Effects on Preferred Viewing Distance**

Chapter 3 introduces the preferred viewing distance task as a first, subjective probe of legibility in VR. Participants adjust a text panel in depth until it feels "clear and comfortable" to read, yielding self-chosen distances that implicitly combine size and clarity under fixed hardware constraints. Using linear mixed-effects models, the chapter examines how these preferences vary with font, glyph complexity, script, and language background. The results show that L1 readers often tolerate smaller angular sizes than L2 readers, consistent with script-tuned top–down compensation and differences in effective perceptual span, and that certain font configurations expand or shrink the comfortable viewing window.

**Chapter 4: Understanding Glyph Characteristics on the Legibility for Native and Non-Native Speakers in Virtual Reality**

Chapter 4 turns to a more constrained, performance-based view of legibility through a glyph identification task. Here, participants see a briefly presented target character in VR and must choose which of two alternatives it matches, under controlled manipulations of glyph complexity and visual similarity. The chapter analyzes error, invisibility, and response time using GLM and LME models to characterize how recognition performance depends on language background and script. The findings reveal that the VR setting preserves deep cross-script differences in visual strategy: character-script natives (Chinese and Japanese) are highly sensitive to similarity and prone to confusion among look-alike forms, whereas alphabetic-script natives are more strongly affected by overall complexity. These patterns clarify which aspects of glyph design are most critical for robust recognition in VR for different reader groups.

**Chapter 5: Study on font preferences of native and non-native speakers in a virtual reality environment**

Chapter 5 returns to a subjective perspective by examining font preference as an aspect of perceived legibility and comfort. Participants view matched text samples rendered in different fonts at fixed positions and sizes in VR, choose which font they find easier or more pleasant to read, and provide brief explanations of their choices. Quantitative analyses of pairwise choices and ratings are complemented by qualitative coding of interview comments. Results show that native readers' preferences tend to track configurations that also support strong performance, whereas L2 readers more frequently favor fonts that are familiar or visually appealing from their own print environments, even when these are not objectively optimal for VR. This chapter highlights how aesthetics, cultural familiarity, and legibility jointly shape user experience with text in immersive media.

**Chapter 6: Discussion**

Chapter 6 provides an integrative discussion of the findings from Chapters 3–5 within an MTMM-inspired framework. It synthesizes convergent patterns across the three methods—such as the role of top–down compensation, perceptual span, and script-specific visual strategies—and identifies cases where performance measures and preferences diverge. The chapter relates these patterns to broader HCI and cognitive frameworks (e.g., perceptual span, cognitive load, and affordances), and develops design-oriented interpretations for multilingual VR interfaces. Building on this synthesis, Chapter 6 also introduces a prototype Visual Text Legibility System (VTLS) implemented in Unity, which encodes the convergent findings as rules and simple predictive functions for legibility-aware text rendering in VR. The chapter concludes by explicitly articulating the main contributions of the thesis, discussing limitations of the current studies (e.g., controlled stimuli, hardware constraints, sample size), and outlining key directions for future work, including extensions to AR, audio–visual integration, and adaptive personalization.

**Chapter 7: Conclusions**

Chapter 7 offers a concise conclusion and broader reflection. It recaps how the unified experimental framework, the MTMM-based comparison of legibility and preference, and the VTLS prototype together provide an initial foundation for systematic, multilingual VR text design. Rather than repeating detailed limitations or methodological points, the final chapter focuses on the overarching implications: that legibility in VR is jointly shaped by device constraints, visual design, and the reader' s linguistic history, and that respecting this interplay is crucial for inclusive, user-centred immersive systems. The chapter closes by situating the thesis within the wider trajectory of immersive computing, arguing that future VR/AR technologies should treat legibility not as a fixed parameter, but as a dynamic, user-sensitive aspect of interaction design.

# Chapter 2 Method

## 2.1 Apparatus

The experiments were conducted on a Meta Quest 3 (Elite Strap) head-mounted display with dual LCDs (2064 × 2208 pixels per eye; 90 Hz). Central angular sampling is approximately 25 pixels/degree. The VR application was built in Unity on a Windows 11 PC (Intel Core i7-13700HX, 32 GB RAM, NVIDIA GeForce RTX 4060 Laptop GPU, 512 GB SSD). For maximum clarity the headset was used tethered via Oculus Link at "Ultra" quality; 8 × MSAA was enabled, text was rendered with TextMesh-Pro signed-distance-field fonts, and dynamic resolution was disabled so that the per-eye render resolution remained fixed across conditions.

Participants interacted with Quest Touch Pro controllers. To reduce postural load, the virtual text panel was head-referenced: its position updated with head motion and remained centered at a preset virtual distance relative to the camera, keeping the content within a comfortable field of view while minimizing fatigue.

## 2.2 VR System Overview

All three empirical chapters share a common VR text legibility system(Fig. 2.1). The same three native-language groups (CL1, JL1, EL1), the same VR environment (Meta Quest 3 with the Unity-based application described above), and the same stimulus library (multilingual language environments, glyph sets, font sets, and four base viewing distances) are used across all modules. By keeping these components fixed, differences in the results can be attributed to the task logic and stimulus manipulations rather than to changes in hardware or context.

On top of this shared platform, the system implements three task modules that correspond to Chapters 3–5. Module 1 (Chapter 3) measures the effect of viewing distance on legibility by asking participants to adjust the text panel until it reaches a comfortable, readable distance. Module 2 (Chapter 4) probes glyph and character structure by requiring participants to identify a target among visually similar alternatives (A vs. B). Module 3 (Chapter 5) focuses on font style, collecting subjective judgments and preferences for different fonts and weights under controlled viewing conditions.

In Chapter 6, the results from these modules are integrated using the Multi-Trait Multi-Method Matrix Method (MTMM) perspective to identify effects that are common across tasks and those that are language-specific. Based on these findings, practical

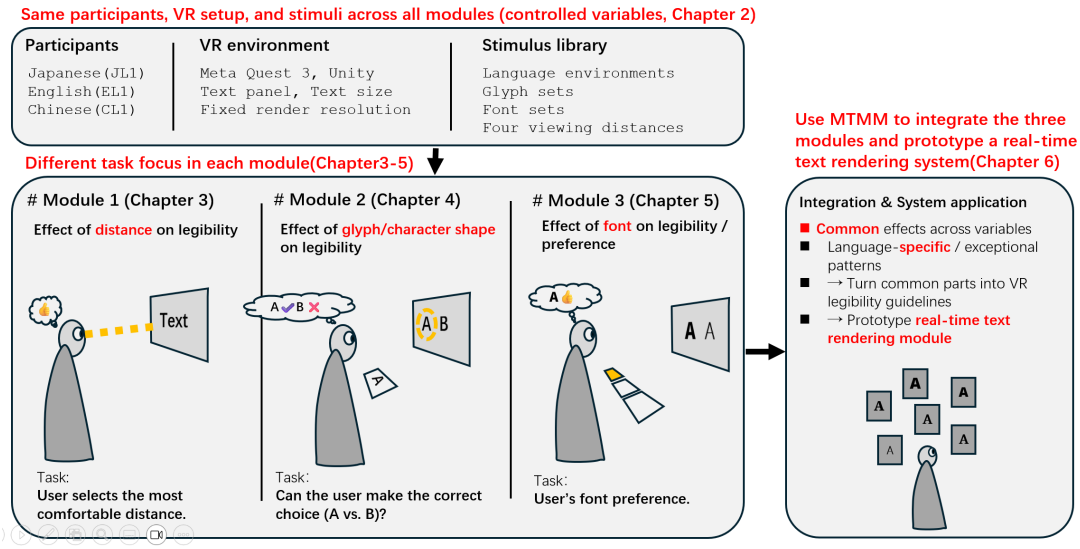



Fig. 2.1: Overview of the VR system

guidelines for VR text rendering are proposed, and a preliminary prototype of a real-time text rendering module is presented to demonstrate how text presentation can adapt to viewing conditions.

## 2.3 Parameters

### 2.3.1 Native and Non-native

In the present study, thirty participants were recruited through flyers and social media. Participants were eligible if they (a) were between 20 and 40 years of age, (b) self-identified as native speakers of Chinese, Japanese, or English, (c) reported regular reading experience in their native script, and (d) had normal or corrected-to-normal vision with no known neurological or visual disorders that would interfere with VR use.

To enable balanced cross-language comparisons, we aimed to recruit ten participants per native group, resulting in 10 native Chinese speakers (CL1; 6 males, 4 females), 10 native Japanese speakers (JL1; 5 males, 5 females), and 10 native English speakers (EL1; 7 males, 3 females).Given the factorial design and the modest sample size ($N = 30$, 10 native readers per language group), we also conducted post-hoc power checks for the GLM models; details are provided in Appendix B. To control for the potential influence of font on different age groups, as highlighted in Calabrese *et al.*'s research on MNREAD Acuity Charts across age ranges (8–16, 16–40, and over 40 years) [47], all participants were between 20 and 40 years of age. Of these, eight had no prior experience with VR headsets, while 22 had used VR headsets before. All participants had normal or corrected-to-normal vision, with 11 having normal vision, two wearing contact lenses, and 17 wearing glasses.

In linguistics and psycholinguistics, L1 denotes the native (first-acquired) language, whereas L2 denotes a non-native language learned after early childhood [48]. However, within L2, the gap between complete beginners and highly proficient users is substantial. Although prior work has sometimes labeled unfamiliar or barely learned target languages as LN, indicating "no prior knowledge" [49], the present study involves three native-language groups and three target languages. Introducing LN as an additional category would therefore yield 27 distinct labels (3 languages $\times$ 3 native groups $\times$ 3 categories), which could be confusing. To preserve readability, only the native-group labels defined at first mention are used: CL1 (native Chinese), JL1 (native Japanese), and EL1 (native English). When these groups read another script, the situation is described in plain terms (e.g., "Japanese native speakers (JL1) reading English text").

To distinguish participants' competence across L1 and L2 (and across different L2s), a language self-assessment was administered to gauge familiarity with each language. Because the focus was on perceptual responses to text rather than linguistic proficiency per se, the items were tailored as follows: for Chinese, familiarity with common Hanzi; for Japanese, (i) familiarity with Kana and (ii) familiarity with common Kanji; and for English, familiarity with words. Responses were given on a four-level scale, ranging from native-like (Expert) to None, as shown in Table 2.1.

Table. 2.1: Self-assessment items with anchor examples

| **Target script** | **Prompt (one choice)** | **Expert** | **Familiar** | **Basic** | **None** |
|---|---|---|---|---|---|
| Chinese (Hanzi) | "How well do you recognize common Chinese characters?" | fluent; rare lookup | most; some lookup | some; freq. lookup | none |
| Japanese (Kanji) | "How well do you recognize Japanese Kanji?" | fluent; rare lookup | many; some lookup | some; freq. lookup | none |
| Japanese (Kana) | "How well do you recognize Japanese Kana (hira-gana/katakana)?" | fluent; rare hes. | most; some hes. | some; slow/errors | none |
| English (words/letters) | "How well do you recognize English words and letters?" | fluent; rare check | most; some check | some; freq. check | none |

*Notes.* One option was selected per row. L1/L2/LN per script: L1 = native (self-reported); L2 = non-native with at least Basic; LN = None ("do not recognize any characters/words").

Based on these items, participants' self-rated familiarity was summarized by native group for each target script, and the resulting distributions are shown in Fig. 2.2. For Japanese (Kana and Kanji) and for English, responses of None are rare among non-native readers (CL1 and EL1 when reading Japanese, and CL1 and JL1 when reading English), which indicates at least some exposure to the L2. In contrast, for Chinese (Hanzi), most JL1 responses fall at Basic (L2), whereas about 70% of EL1 responses fall

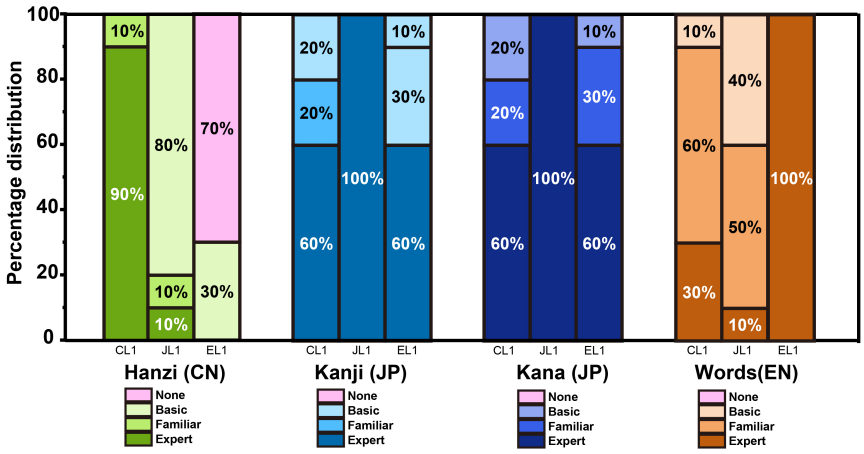



Fig. 2.2: Self-assessed familiarity distributions by native group (CL1/JL1/EL1) across four target scripts (Hanzi, Kanji, Kana, English words).

at None. Accordingly, when discussing EL1 reading Chinese, this group is treated as near-LN or, at most, a very low-level L2 group.

### 2.3.2 Distance

Most VR headsets render all content at a single, fixed focus distance; the Meta Quest 3 is of this type. As a result, simulated depths that are very near or very far tend to be less comfortable and less legible than mid-range depths [50], as illustrated in Fig. 2.3. Alger et al. also emphasize that, in VR, information should not be placed closer than 0.5 m or farther than 20 m to ensure optimal legibility [51]. Guided by these findings, the virtual environment in this study sampled four representative viewing distances: 0.5 m , 2.5 m, 5 m, and 10 m. Because the mid-range is relatively broad, it was divided into a comfortable-near condition (2.5 m) and a comfortable mid-far condition (5 m), while 0.5 m and 10 m serve as the near and far limits, respectively. Operationally in the experimental tasks, these distances span: (0.5 m) the nearest distance at which characters remained readable, (2.5 m) a comfortable near distance, (5 m) a comfortable mid-far distance, and (10 m) a far limit where character contours approach the threshold of recognizability.

### 2.3.3 Glyph

#### Characteristics of glyphs in different languages

To enable cross-script comparisons grounded in each writing system' s visual organization, this study treats the glyph as the minimal perceptual unit and defines it for each language in accordance with its orthographic structure. This subsection summarizes the

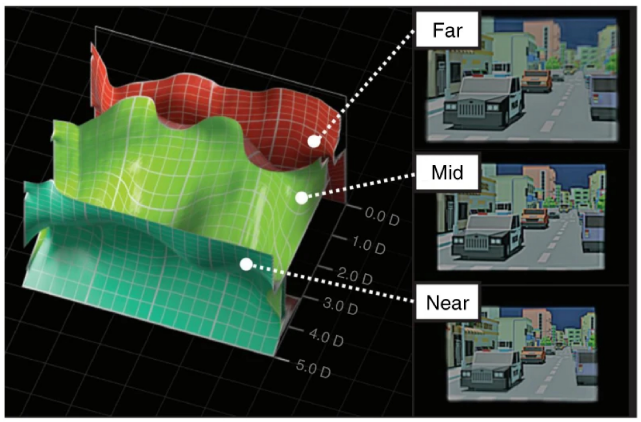



Fig. 2.3: Near–mid–far depth zones in near-eye VR displays (conceptual).[50]

script-internal features used for stimulus construction and analysis.

**Chinese (logographic).** Chinese encodes meaning within compact, two-dimensional character forms. A single Hanzi combines multiple subcomponents within a bounded square, yielding high within-character density. For example, the character 赢 (yíng, "to win") integrates five radicals (亡、口、月、贝、凡) into a cohesive configuration. Such spatial integration reflects a logographic strategy in which each character is a visually self-contained unit with semantic value [30]. In the present framework, **glyphs are defined as individual Hanzi**, each functioning as a unified visual unit.

**Japanese (logographic–syllabic).** Japanese combines two systems: logographic Kanji and syllabic Kana. Kana were historically derived from cursive simplifications of Chinese characters during the Heian period (794–1185 CE), forming two syllabaries (Hiragana and Katakana) that are visually compact and discrete [52]. Although Kana encode syllables rather than morphemes, their letterforms constitute clear, self-contained visual chunks, and L2 learners may perceive Kana in a gestalt-like manner similar to Kanji due to visual chunking [53]. In the present framework, **glyphs include both Kanji and Kana**, treated uniformly as discrete perceptual units.

**English (alphabetic).** English distributes information across sequences of letters (graphemes) arranged linearly, so visual processing proceeds over ordered elements rather than within a single square form. For example, the word *unbreakable* extends horizontally by combining the prefix *un-*, the root *break*, and the suffix *-able*. This sequential organization contrasts with the spatial integration of Chinese characters and places a premium on serial decoding of letter strings [30, 54]. In the present framework, because single letters carry no standalone lexical meaning, **glyphs are operationalized as whole words**, reflecting the smallest meaningful and task-relevant visual unit in English.

To ensure cross-linguistic consistency in the experimental design, glyphs are defined in visually oriented terms as the minimal perceptual units in each writing system. Under this definition, a glyph may correspond to (a) a single logographic grapheme (e.g., a Chinese Hanzi character such as 水 shuǐ, "water"), (b) a single syllabic sign (e.g., a Japanese kana such as "き"), or (c) a composed glyph combining multiple alphabetic graphemes (e.g., an English word such as "experiment"). This framework allows systematic comparison across languages:

- Chinese: **Glyphs are defined as individual Hanzi** (e.g., 贏 ), each representing a unified visual unit.
- Japanese: **Glyphs include both Kanji** (logographic characters) **and Kana** (syllabic symbols), treated as discrete perceptual chunks.
- English: **Glyphs are operationalized as whole words**, reflecting their linear composition from alphabetic graphemes.

In other words, throughout this dissertation the notion of a "glyph" is deliberately defined in visual–perceptual terms. Across all subsequent experiments, glyphs are used as controlled visual tokens rather than as carriers of sentence- or discourse-level meaning, and the tasks are designed such that successful performance does not require understanding the lexical content of the stimuli. As a result, relationships among word or sentence meanings across Chinese, Japanese, and English are not analyzed here; instead, the studies focus on how readers perceive and prefer script-specific visual forms (e.g., stroke complexity, configuration, and layout) under VR viewing conditions. How these perceptual findings interact with semantic or contextual processing is left for future work that explicitly incorporates reading-comprehension tasks.

To illustrate these structural differences, a comparative diagram was prepared, as shown in Figure 2.4. This figure illustrates the differences in visual structure among Chinese (logographic), Japanese (logographic–syllabic), and English (alphabetic) writing systems. Chinese characters combine multiple components into a compact two-dimensional form (e.g., the character 使 ), exemplifying "structural complexity." In contrast, English words extend along a horizontal axis by adding letters in sequence, reflecting "sequential complexity." Japanese incorporates both kanji and kana, offering a blend of visual and linear elements. In this system, kana and individual English letters function as basic units without inherent meaning, whereas a single Chinese character or a complete English word serves as the smallest meaningful unit. The concept of a "glyph" is introduced here as a unifying framework for cross-linguistic comparison.

### Glyph Complexity

To ground the glyph stimuli in each language' s writing system, the smallest meaningful or visually discrete units available in each script were selected. In logographic

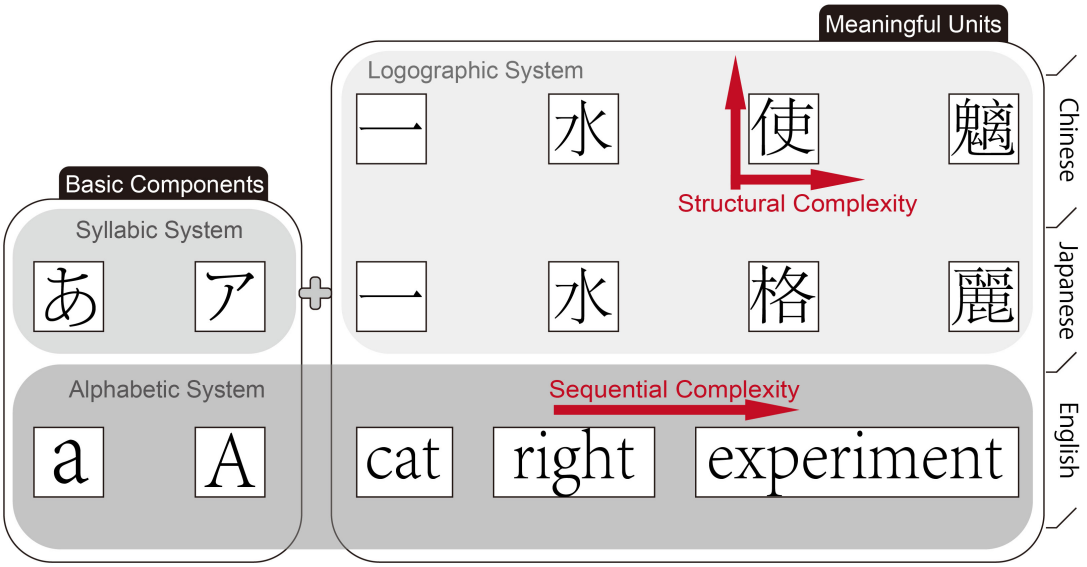



Fig. 2.4: A Cross-Linguistic Comparison of Orthographic Structures.

and mixed-script languages (Chinese and Japanese), individual characters (Hanzi, hiragana, katakana, and Kanji) correspond directly to morphemic or syllabic units and remain visually self-contained. Accordingly, the Chinese and Japanese conditions presented isolated characters (following Wang (2020) [55] and Zhou (2024) [56]), whereas the English condition presented isolated high-frequency words. Participants were not asked to perform any reading or comprehension task; instead, they made purely perceptual legibility judgments. This design ensures that the results primarily reflect visual distinctions between fonts rather than lexical or semantic differences.

Based on the characteristics of each language, the fonts were grouped accordingly. For this experiment, the glyphs were categorized into three main sets: kana, Chinese characters (Hanzi and Kanji), and the Latin alphabet, corresponding to Japanese, Chinese, and English, respectively. Chinese characters (Hanzi and Kanji) are visually block-like, whereas Latin letters consist of linear strokes arranged along a baseline.

**Chinese Character(Hanzi and Kanji) Glyphs** Because perceived complexity varies within each script, glyphs were classified according to their visual complexity. For block-like characters (e.g., Chinese and Japanese characters), complexity arises primarily from the number of strokes, with a larger stroke count typically indicating greater complexity. Various methods have been proposed to quantify the complexity of Chinese characters, including stroke count, skeleton length (total stroke length) as discussed by Bernard and Chung (2011) [57], the slice method (calculating the frequency of strokes in horizontal, vertical, and diagonal slices and taking the maximum intersection number) [58], and the square of the symbol’ s perimeter divided by the “ink” area [59]. In the present exper-

iment, the complexity of both Japanese and Chinese characters was quantified with a composite measure based on stroke count and ink area. For each character, the number of strokes was obtained from standard orthographic descriptions, and the character was rendered at a fixed size to estimate ink area as the proportion of foreground ("ink") pixels in the glyph bitmap. These two quantities were rescaled to a common range and combined into a single index, yielding a composite measure in which higher values correspond to characters that are both more stroke-dense and visually heavier (occupying more inked area). In this way, the measure balances the structural information captured by stroke count with the overall visual mass reflected in ink area, consistent with prior work emphasizing both contributions to perceived complexity.

Considering the subtle differences between Chinese Hanzi and Japanese Kanji, the Kanji materials used in this experiment were sourced from a commonly referenced Japanese Kanji list based on the Cabinet notification of November 30, 2010 [60], while the Hanzi were sourced from a general-standard Chinese character dataset hosted on GitHub [61]. For each character, the stroke count was determined and the pixel value at 100 pt was calculated using the Pillow image-processing library in Python [62]. The complexity of each character was then calculated using the following formula:

$$C = \frac{NUMBER}{\mu number} \times \frac{PIXEL}{\mu pixels} \tag{2.1}$$

where "NUMBER" represents the stroke count, "PIXEL" is the pixel value, and $\mu_{number}$ and $\mu_{pixels}$ are the average stroke count and pixel values, respectively.Figure 2.5 illustrates these two marginal distributions using Japanese Kanji as an example; the vertical arrows mark the sample means used for normalization. Although characters with more strokes tend to be more complex, stroke count alone treats visually tiny strokes (e.g., the dot "丶" in 文) the same as long strokes (e.g., the horizontal "一" in 文). By incorporating the pixel footprint, the measure captures both structural and visual complexity.

The resulting complexity distribution $C$ is strongly right-skewed (median $\tilde{C} \approx 2.79$ for Chinese, $\tilde{C} \approx 2.27$ for Japanese; 75th percentile $C_{0.75} \approx 3.66$ for Chinese, $C_{0.75} \approx 2.99$ for Japanese). To mitigate this skew while retaining intuitive, data-informed tiers, $\log_2 C$ was discretized at cut-points $-1$, 1, 2 (equivalently, thresholds on the original scale $C \in \{2^{-1}, 2^1, 2^2\} = \{0.5, 2, 4\}$), yielding four levels: $C < 0.5$, $0.5 \leq C < 2$, $2 \leq C < 4$, and $C \geq 4$.

**Glyph complexity of Japanese Kana** From a glyph–shape standpoint, Kana are also block-like characters confined to a square em box—even though they are syllabic —and historically derive from simplified Chinese letterforms [63]. Applying the composite complexity index $C$ defined in Section 2.3.3 is applied, kana cluster at the bottom of the distribution: in the sampled set, individual hiragana/katakana glyphs contain no more than four strokes, with small pixel footprints and few internal intersections. Consequently, kana were assigned to the lowest complexity tier ("very simple",

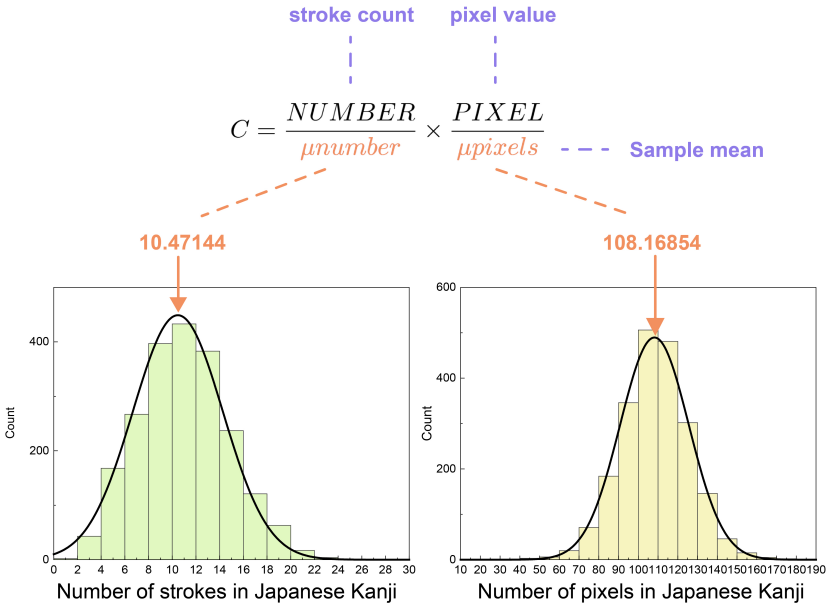



Fig. 2.5: Schematic illustration of the complexity measure using Japanese Kanji.

$C \leq 0.5$),placing them on the same $C$-scale as the simple Hanzi/Kanji tier defined by the identical threshold.

**Glyph complexity of English words** In alphabetic English, single letters are graphemes without standalone meaning, whereas single words are the minimal meaningful lexemes. For the linear-script condition (English), words were grouped into three length-based categories: short (up to 3 characters), medium (4–8 characters), and long (9 or more characters). This classification method aligns with the approach used by Jukka [64] in their study on complexity categorization for Finnish, an alphabetic system.

### 2.3.4 Font: font style and font weight

Extant research has established that font selection selection exerts a critical influence on text legibility and cognitive processing in virtual reality environments [65–69]. Moreover, a wealth of empirical evidence shows that, to achieve equivalent reading performance with Latin letters, CJK scripts typically require larger nominal font sizes [41, 70, 71], and this insight has been integrated into font design. As Figure 2.6(a) illustrates, at the same point size Latin letters occupy less area and appear visually light; Chinese characters—with dense, high-contrast strokes—occupy more visual area and appear heavier and hiragana and Kana, occupy the same nominal dimensions but have slightly lower visual fill than Chinese characters, giving them a comparatively softer, more dispersed appearance.

As shown in Figure 2.6(b), typefaces are generally divided into two main categories:

**sans-serif**, which feature consistent stroke widths and almost no ornamentation—making them ideal for on-screen body text—and **serif**, which themselves split into two traditions: one that preserves the restrained proportions of traditional print (with relatively fine horizontal strokes) and another that draws on calligraphic precedents (with deliberate stroke-thickness variation reminiscent of brushwork). In either case, serif faces are defined by decorative terminals—small "feet" or flares at the beginning and end of strokes. Beyond font style, every typeface is commonly offered in multiple weights (Thin, Medium, bold), and adjusting stroke thickness in this way further modulates readability. The combination of font style (serif vs. sans-serif) and weight produces distinct visual textures and legibility profiles across different display contexts.

Across all studies in this dissertation, text stimuli were presented in four writing systems: the Latin alphabet, Chinese logographs (Hanzi/Kanji), Japanese Hiragana, and Japanese Katakana. To minimize design-related confounds and ensure consistent coverage across scripts, harmonized variants from the Google Noto superfamily were employed [72–77].

Typeface was manipulated along two orthogonal factors in a 2 (style) × 3 (weight) design:

- **Font style**: Sans-serif vs. Serif
- **Font weight**: Thin, Medium, Bold

Crossing these levels yields six style×weight combinations per script. Unless otherwise noted, stimuli in Chapters 3–5 were rendered at matched nominal point sizes with the fonts' default metrics (letter-spacing/kerning and hinting), so that observed effects can be attributed to typeface style and weight rather than ad-hoc adjustments. Figure 2.6(c) illustrates the six combinations using the word "apple" (or script-appropriate forms) as a neutral specimen.

## 2.4 Method

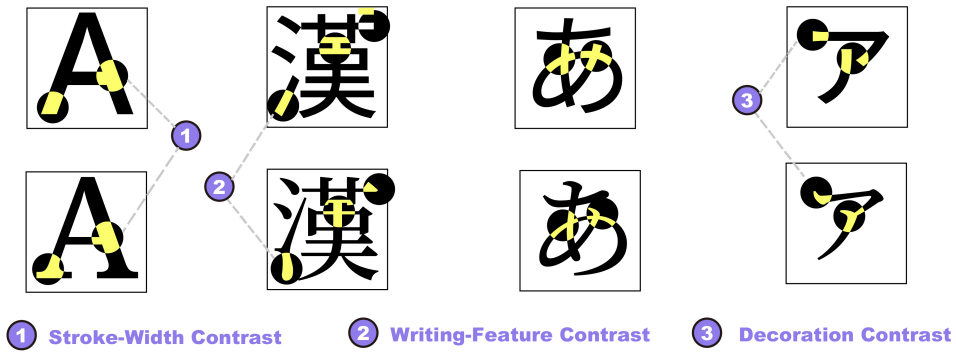
This research investigates how viewing distance, glyph (character) structure and complexity, and typeface features interact with native (L1) versus non-native (L2) language background to shape character identification and perceived visual clarity in VR. The overarching goal is to isolate basic perceptual differences in VR text perception rather than language comprehension per se.

To increase the reliability and generalizability of the conclusions, three independent yet complementary experiments were conducted. The overall design is inspired by the logic of converging operations and the multitrait–multimethod (MTMM) approach to construct validation [78]. In classic MTMM work, multiple traits (e.g., depression vs. anxiety) are each measured by multiple methods (e.g., self-report questionnaire vs. clinician rating), and a matrix of correlations is inspected to separate what is shared at the trait level from what is specific to the measurement method [78, 79]. When different methods that are

(a)




(b)




(c)

| English | Chinese | Japanese | | | Font | |
|---|---|---|---|---|---|---|
| Word | Hanzi | Kanji | Katakana | Hiragana | Font Name | Abbreviation |
| apple | 苹果 | 林檎 | リンゴ | りんご | Sans-Thin | SaT |
| apple | 苹果 | 林檎 | リンゴ | りんご | Sans-Medium | SaM |
| **apple** | **苹果** | **林檎** | **リンゴ** | **りんご** | **Sans-Black** | **SaB** |
| apple | 苹果 | 林檎 | リンゴ | りんご | Serif-Extralight /Serif-Thin | SeE /SeT |
| apple | 苹果 | 林檎 | リンゴ | りんご | Serif-Medium | SeM |
| **apple** | **苹果** | **林檎** | **リンゴ** | **りんご** | **Serif-Black** | **SeB** |

Fig. 2.6: Typeface Overview Diagram. (a) Comparison of area occupied by four writing systems at the same point size. (b) Examples of sans-serif versus serif font styles. (c) Sample of six fonts combinations: 2 styles (Sans, Serif) × 3 weights (Thin, Medium, Bold) across four scripts.

intended to tap the same trait converge, and when conceptually distinct traits remain discriminable, confidence in the underlying constructs increases.

In the present thesis, this MTMM logic is adopted at the design level rather than through a full CFA–MTMM statistical model. Two broad traits are of interest: (i) *perceptual legibility* (how easily text can be seen and identified), and (ii) *subjective preference/comfort* (how fonts and layouts are experienced by users). These traits are probed by three methods that share a common set of visual parameters (viewing distance, glyph complexity, font features, and L1/L2 grouping): a preferred viewing distance adjustment task (Chapter 3), a forced-choice glyph identification task (Chapter 4), and a subjective font rating task (Chapter 5). Table 2.2 summarizes this multitrait–multimethod structure and its mapping onto the experimental chapters.

In parallel, the thesis adopts a mixed-methods stance common in HCI. Quantitative indices (e.g., accuracy, invisibility rates, response time, and preferred distance) are complemented by qualitative evidence (e.g., comfort-based distance justifications, font-

Table. 2.2: Schematic MTMM layout of traits and methods used in the thesis.

| **Traits** | **Methods** | | |
|---|---|---|---|
| | Preferred viewing distance (Chap. 3) | Glyph identification (Chap. 4) | Font preference (Chap. 5) |
| Legibility | Chosen distance at near-threshold clarity | Error, invisibility, Task completion time | Clarity ratings |
| Preference | Comfort-based distance choice across visual parameter settings | – | Overall preference, aesthetics, interview comments |

preference explanations). Triangulating performance data with user experience helps ensure that the findings are both statistically robust and design-actionable [80, 81].

Operationally, each experiment targets a distinct facet of VR text perception—preferred viewing distance, glyph-level legibility under complexity and similarity, and font preference under controlled viewing—while sharing the common parameter set defined above. This MTMM-style structure makes it possible to test whether L1–L2 differences replicate across methods (convergent evidence for trait-level effects) and to identify aspects of performance or preference that are method-specific (discriminant evidence), thereby aligning subjective and objective evidence within a single comparative framework.

# Chapter 3 Perceiving Multilingual Text in Virtual Reality: Glyph Complexity and Font Effects on Preferred Viewing Distance

This chapter takes preferred viewing distance as its central outcome and asks how it is shaped jointly by VR-specific perceptual constraints and basic typographic structure. Text readability in head-mounted displays depends not only on retinal (angular) size but also on perceived reading distance. A near, phone-like UI and a far, cinema-like UI can be matched in visual angle and still feel different because accommodation–vergence demands, motion parallax, spatial context, and size-constancy mechanisms reshape perceived scale, effort, and comfort [82]. This matters in VR, where limited display resolution, variable viewing distances, and depth-of-focus demands already slow reading and increase visual strain compared to paper or standard screens [83, 84]. As a result, typographic and visual choices—serif versus sans-serif, stroke weight, and glyph complexity—become especially consequential [65, 66]; thin strokes degrade under motion [67], whereas bolder, high-contrast designs improve legibility and reduce visual effort [68, 69].

In multilingual VR use, script differences further modulate distance needs: Chinese Hanzi and Japanese Kanji have higher within-character stroke density than Latin letters, Kana differ in curvilinearity vs. rectilinearity [63], and Chinese newspaper text requires about 1.5× larger character size (arcminutes) than English for equivalent legibility [85]. Meanwhile, headset field of view (FoV) and platform distance-independent angular units (e.g., dmm) are used in practice to stabilize text sizing across distances and devices [86–88]. Together, these points motivate treating preferred viewing distance as a first-class design and measurement target, complementary to angular-size specifications.

## 3.1 Related Work

### 3.1.1 Psychophysics of distance perception and comfort in VR

A robust literature shows that egocentric distances tend to be underestimated in virtual environments relative to the real world; the bias is moderated but not eliminated by

display and task factors [82]. In head-mounted displays, the fixed-focus optics decouple vergence and accommodation, producing viewing conditions that can constrain comfort even when angular size is sufficient [89]. FoV further shapes perceived scale and distance judgements [90], implying that two text blocks matched in visual angle can differ in subjective workload and usability at different absolute distances. These psychophysical factors provide the foundation for analyzing distance-dependent readability in VR beyond angular-threshold rules.

### 3.1.2 Design factors that influence distance-dependent readability in VR

Low-level text properties and viewing parameters jointly determine legibility in immersive displays. Character complexity (indexed by stroke count) raises recognition thresholds for Chinese characters, requiring larger angular sizes or nearer distances [85]. Font features—presence of serifs, stroke weight, x-height—systematically modulate readability as distance and size change [88, 91–93]. Serifs can aid contour following at large sizes or short distances but fragment at small angular sizes under limited resolution, reducing legibility [91, 92]. While serif vs. sans-serif findings are mixed on conventional 2D displays [94], VR work consistently shows that sans-serif and heavier weights support reading at greater distances by counteracting fading of fine strokes [88, 93]. Head or text motion penalizes thin strokes [67], whereas bold, high-contrast settings reduce visual effort [68, 69].

Beyond typography, platform-facing guidance advocates specifying text by minimum angular extent so that size remains stable across distances and devices; Google popularized distance-independent millimeters (dmm) for VR text at Google I/O 2017, and engineering recommendations specify minimum angular extents for body text at typical viewing distances [86–88]. FoV and resolution interact with these specifications to set the practical legibility window in which text remains comfortable at a given distance [90]. These design factors jointly motivate coupling distance choices with font and glyph design when placing text in 3D space.

### 3.1.3 Population differences in perceived distance to text

Readers adapt visual strategies with language background and script complexity [95, 96]. Here, L1 denotes the native language and L2 a later-acquired language [48]. Cross-linguistic eye-tracking shows that native Chinese readers processing Japanese develop distinct vertical/horizontal reading behaviors and often position closer to enhance Kana clarity relative to Kanji [97]. The same readers exhibit more frequent and longer fixations and higher regressions when reading English than native English readers, indicating greater processing load [98]. Native Japanese readers compensate for Kanji' s complexity by fixating more on high-stroke characters than on simpler Kana [99], and show a reduced

perceptual span in English compared to native English readers [100]. English L1 readers tackling Chinese benefit from increased spacing that reduces search time and errors [101]; bilingual corpora show L2 logographic reading involves longer fixations, more regressions, and shorter saccades, often offset by closer viewing and repeated previews to secure accurate identification [102]. These patterns suggest that preferred reading distance functions as a compensatory lever varying with script and language background, above and beyond angular-size sufficiency.

Taken together, the literature indicates that angular-size rules alone are insufficient to guarantee comfortable reading in VR. Psychophysical evidence shows that egocentric distances are systematically underestimated in virtual environments and that the vergence–accommodation conflict constrains comfort even when text meets angular-size thresholds [82, 89, 103]. Field of view and resolution further shape perceived scale and the usable range for text, while platform guidance (e.g., distance-independent angular units such as dmm) stabilizes specification but not necessarily experience across distances and devices [86–88, 90]. At the same time, design factors—glyph complexity, serif versus sans-serif style, and stroke weight—interact with distance and size to set recognition thresholds and visual effort [67–69, 85, 91–93]. Finally, readers' language backgrounds and script familiarity (L1 vs. L2; logographic vs. alphabetic) modulate viewing strategies and thus the preferred distance at which text feels clear and effortless [95, 98–102]. Therefore, a distance-aware approach that integrates psychophysical constraints, VR design parameters, and cross-linguistic differences is needed to characterize and model preferred viewing distance beyond angular sufficiency, and to derive adaptive guidelines for multilingual text in VR.

## 3.2 Experiment Design

A controlled VR experiment was conducted using a Meta Quest 3 headset, in which participants from three language groups (Chinese, English, and Japanese) adjusted the distance of text panels, varying in glyph complexity and font style, until a comfortable reading position was reached. By analyzing the preferences and behaviors of multilingual users, the study aims to derive design implications for VR systems. In particular, the findings are intended to inform adaptive text rendering techniques that dynamically adjust text size, distance, or font based on the user's language profile and the content's visual complexity. Such adaptability has the potential to improve accessibility and comfort in VR, ensuring that users can easily read text in both native and non-native languages, regardless of script or font.

To ensure consistency of test conditions, text height was fixed at 70 mm. To address resolution consistency and preserve visual quality, dynamic resolution scaling was enabled. Dynamic resolution scaling [104] adjusted the render resolution in real time based on GPU load, balancing performance and visual sharpness so that clarity was maintained under varying conditions. Prior to data collection, pilot tests were conducted to verify

that text legibility at all distances met the experiment' s visual requirements.

A 14 × 2.5 × 2.5 m virtual corridor environment was created, in which stimuli were presented in three languages (Chinese, Japanese, and English) within a 570 × 200 mm virtual text panel. The panel background was set to black at 50% opacity and text was rendered in white, with each string filling the panel. A prompt box positioned below the user' s field of view randomly indicated which language to attend to (Chinese, Japanese, or English) and instructed participants to adopt the corresponding viewing distance. Directional lighting ensured uniform illumination throughout the scene. A schematic of the setup is shown in Figure 3.1.

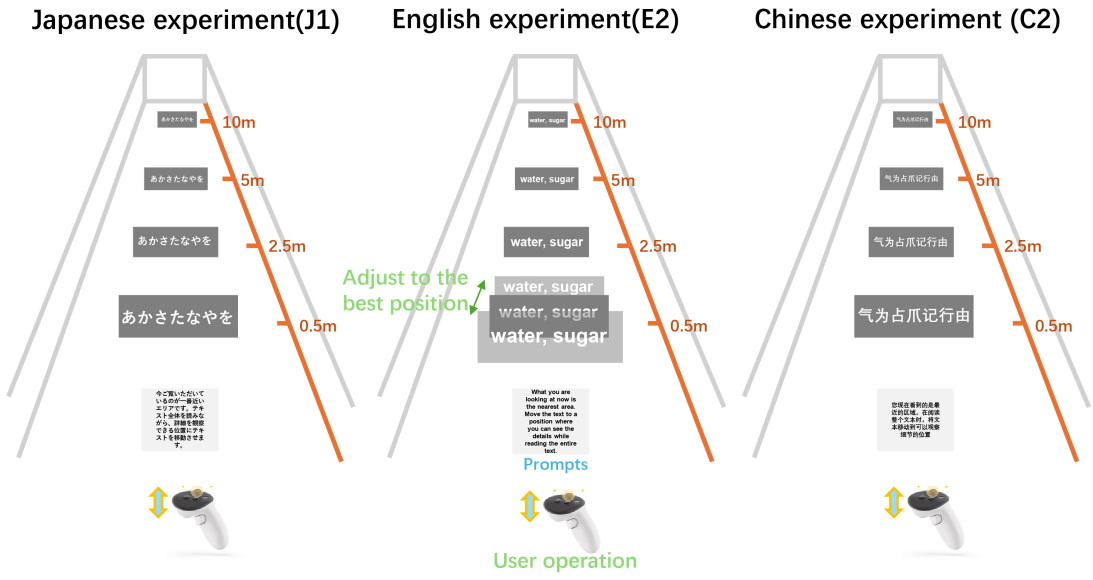



Fig. 3.1: Schematic Diagram of Experimental Setups for Chinese, Japanese, and English Tests.

### 3.2.1 Independent variable

#### Native language

For the native language, following the recruitment of the native language as described in Chapter 2, this experiment also used 10 native Chinese speakers, 10 native English speakers, and 10 native Japanese speakers. The participants were also assigned to the native language and non-native language projects. Therefore, the final participants in this Chinese experiment (10 native Chinese speakers, 5 native Japanese speakers, 5 native English speakers), the participants in the Japanese experiment (10 native Japanese speakers, 5 native Chinese speakers, 5 native English speakers), and the participants in the English experiment (10 native English speakers, 5 native Chinese speakers, 5 native Japanese speakers)

### Distance

Four viewing distances introduced in Section 2.3.2 were used (Figure 3.1): 0.5 m, 2.5 m, 5 m, and 10 m. These correspond, respectively, to (0.5 m) the minimum distance at which characters remain legible, (2.5 m) the comfortable reading distance for myopic observers, (5 m) the appropriate viewing distance for hyperopic observers, and (10 m) the maximal distance at which character contours become nearly indistinguishable. In this chapter, distance is the primary manipulated factor while font style and glyph complexity are crossed as secondary factors; all other task parameters (e.g., target angular sizes and panel implementation in world space) follow the definitions established in Section 2.3.2.

### Font

Font factors followed those defined in Section 2.3.4 (see exemplars in Figure 2.6c). Stimuli were rendered across four writing systems—the Latin alphabet, Chinese logographs (Hanzi/Kanji), Japanese Hiragana, and Japanese Katakana—using harmonized variants from the Google Noto superfamily [72–77]. Typeface is manipulated in a 2 (style: Sans-serif, Serif) $\times$ 3 (weight: Thin, Medium, Bold) design, yielding six combinations per script. Noto was chosen for its broad language support and consistent metrics across scripts, reducing confounds from mismatched design conventions.

### Glyph

To focus on visual recognizability rather than semantics, different minimal units were used for each script. For Chinese and Japanese, 7-character pseudo-texts composed of isolated Hanzi or kana were presented. Such pseudo-texts suppress lexical access and minimize semantic processing, allowing perceptual factors to be probed more directly [55, 105, 106]. For English, the smallest meaningful unit is a word, so high-frequency single words (e.g. “cat”, “water”) were selected to approximate typical reading items while keeping semantic load low.

For Chinese Hanzi and Japanese Kanji, the composite complexity index $C$ defined in Section 2.3.3is used (see Eq. (1) and Figure 2.5 there). Following the same thresholds as in Chapter 2, the four tiers are:

$$C < 0.5, \quad 0.5 \leq C < 2, \quad 2 \leq C < 4, \quad C \geq 4.$$

The index $C$ is not re-derived here; Chapter 2 remains the source of record for computation details and data sources.

Table 3.1 lists the categories used for stimulus sampling across the three scripts:

- **Japanese.** J1 = Hiragana (very simple glyphs, typically $\leq$4 strokes); J2 = Katakana (very simple); J3–J6 = Kanji grouped by $C$ into four tiers: J3 ($C \leq 0.5$), J4 ($0.5 < C \leq 2$), J5 ($2 < C \leq 4$), J6 ($C > 4$). Although Kana are syllabaries

Table. 3.1: Character–glyph classification across Chinese, Japanese, and English. CJK tiers reuse the complexity thresholds from Section 2.3.3.

| Language | Feature | Glyphs | Classification |
|---|---|---|---|
| Japanese | Hiragana (J1) | あかさたなやを | Very simple |
| | Katakana (J2) | アカサタナヤヨ | Very simple |
| | Kanji ($C \leq 0.5$) (J3) | 人一刀才七夕了 | Simple |
| | Kanji ($0.5 < C \leq 2$) (J4) | 爪切匂収企家柔 | Normal |
| | Kanji ($2 < C \leq 4$) (J5) | 辞晴晩娘魅戦富 | Complex |
| | Kanji ($C > 4$) (J6) | 機臓襲職艦騰驚 | Very complex |
| Chinese | Hanzi ($C \leq 0.5$) (C1) | 卜乙二一八入厂 | Simple |
| | Hanzi ($0.5 < C \leq 2$) (C2) | 气为占爪记行由 | Normal |
| | Hanzi ($2 < C \leq 4$) (C3) | 丞泌瓮鸠烫涉逊 | Complex |
| | Hanzi ($C > 4$) (C4) | 瀛鹦蠱饕僭颧蠣 | Very complex |
| English | 1–3 letters (E1) | cat, dog, sun | Simple |
| | 4–8 letters (E2) | water, sugar | Normal |
| | >8 letters (E3) | responsibility | Complex |

linguistically, their letterforms derive historically from Chinese characters and are visually compact [63]; Consequently, J1/J2 are treated as the "very simple" end of the Japanese complexity spectrum.

- **Chinese.** C1–C4 = Hanzi grouped by $C$ into the same four tiers as Kanji: C1 ($C \leq 0.5$), C2 ($0.5 < C \leq 2$), C3 ($2 < C \leq 4$), C4 ($C > 4$).
- **English.** Because stroke-based metrics are not meaningful for alphabetic words, word length was used as a proxy for visual span: E1 (1–3 letters), E2 (4–8 letters), E3 (>8 letters). This keeps the English materials comparable to CJK in terms of increasing visual load with category.

As in Chapter 2, the analysis focuses on visual recognizability rather than semantics; script-specific categories therefore reflect visual demands—stroke density and spatial coverage for Chinese and Japanese characters, and word length (visual span) for English —rather than the complexity of language comprehension [57–59].

### 3.2.2 Task

Before the experiment began, participants were briefed on the perceptual implications of four viewing distances:

- 0.5 m: An extreme near distance at which text appears very large. Participants were instructed to move the text panel forward or backward until it reached the closest position they could comfortably read without moving their head or straining their eyes.
- 2.5 m: A comfortable "near - vision" distance, analogous to typical reading distance. Participants adjusted the panel until it felt like a natural reading distance.

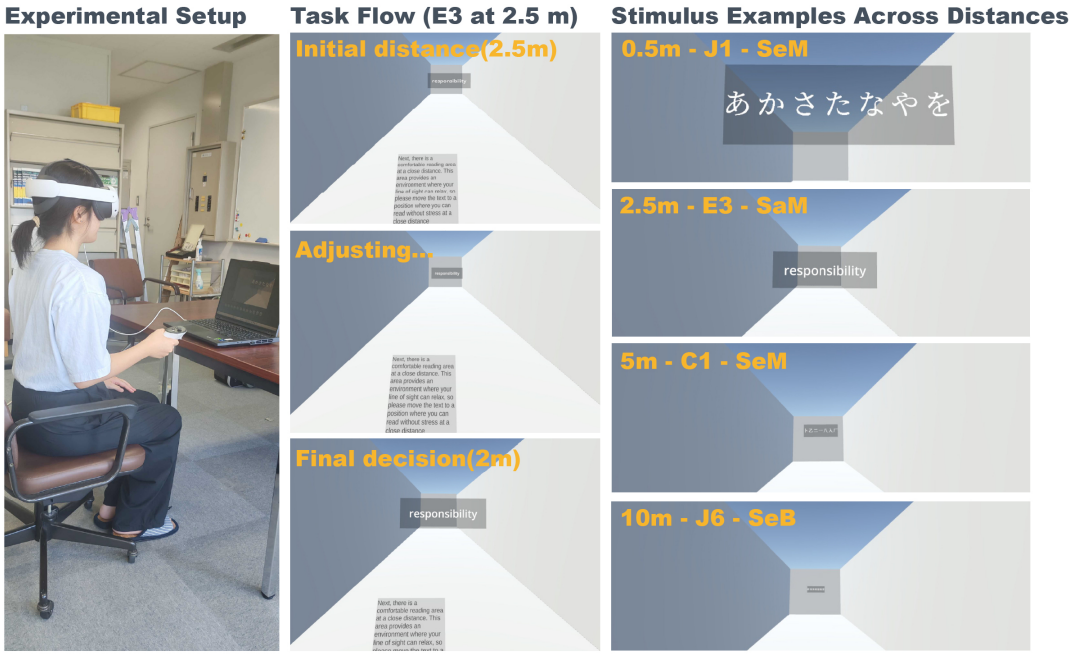



Fig. 3.2: Overview of the Distance-Adjustment Task.

- 5 m: A comfortable "far - vision" distance, akin to reading a distant billboard. Participants again moved the panel until it felt comfortable.
- 10 m: An extreme far distance at which characters begin to blur. Participants were asked to adjust the panel until the text was just barely recognizable.

Each trial began with the text panel positioned at one of these four distances (0.5, 2.5, 5, or 10 m). Participants then adjusted the panel along the depth axis to meet the condition defined for that distance. Once they confirmed their adjustment, the panel was reset to its initial distance for the next trial. All font - style and character - set combinations were completed at one distance before moving on to the next. To counterbalance order effects, half of the participants experienced the distances in ascending order (0.5 m $\rightarrow$ 10 m) and half in descending order (10 m $\rightarrow$ 0.5 m). Participants first completed the task in their native language, took a short break, and then repeated it in a randomly assigned non - native language. Each language block lasted approximately 10–15 minutes. Figure 3.2 summarizes the in-lab setup, the adjustment flow (shown for E3 at 2.5 m), and stimulus examples at the four starting distances across scripts and fonts.

### 3.2.3 Procedure

Participants were first briefed on the study' s objectives and procedures, and informed consent was obtained. After signing consent and withdrawal forms and complet-

ing a baseline questionnaire, they proceeded to VR training, which involved donning the headset, adjusting the device, and learning to operate the Meta Quest Touch Plus controllers. A short pilot session was then conducted to ensure that each participant could successfully complete the experimental task.

Each participant completed two reading-distance calibration tasks: one in their native language (L1) and one in a randomly assigned non-native language (L2). For each language condition, 10 native-language speakers and 10 non-native speakers (with the remaining two language groups comprising five participants each) were evenly assigned to one of two distance-order sequences: ascending (0.5 m $\rightarrow$ 10 m) or descending (10 m $\rightarrow$ 0.5 m).

In the main experiment, participants adjusted the text panel to the appropriate position for each of the four distances (0.5, 2.5, 5, and 10 m) using stimuli that varied by character-set complexity and font style. Upon completing the first language block, they took a short break before performing the same procedure in the second language. Finally, participants completed a post-experiment questionnaire to rate subjective comfort and provide feedback. Total participation time did not exceed 30 min.

### 3.2.4 Data collection and analysis

Based on participants' baseline questionnaires, glyph, font, and initial viewing distance were treated as independent variables, and the participant-adjusted final distance served as the dependent variable for each trial. To obtain a more robust outcome measure, a distance-difference score was computed:

$$FDdeID = \text{Final distance} - \text{Initial distance},$$

such that positive values indicate that the panel was placed farther than the initial distance, whereas negative values indicate a closer placement.

Participants' text distance adjustments were analyzed using a linear mixed-effects (LME) model fitted separately for each language condition (Japanese, Chinese, and English text). The model included fixed factors for font (FontName), character complexity (GraphComplexity), reader group (native vs. non-native; NativeLanguage), and their interactions, with participant ID entered as a random factor. To handle unbalanced data, sum-to-zero coding was adopted for categorical factors, so that effects are expressed relative to the grand mean and the resulting $p$-values are more conservative and interpretable. Estimated marginal means (EMMs) were then derived from the model to interpret the predicted viewing distances under each condition. All analyses were implemented in Python [107], with data handling and prediction grids carried out using pandas [108] and NumPy [109].

The dependent measure was the distance-difference score, defined as the final distance minus the initial distance ($FDdeID$) set by the participant. A positive $FDdeID$ indicates that the text was placed farther away than the initial position, whereas a negative

value indicates that it was brought closer. The following sections summarize the key effects for each text language at each viewing distance from 0.5 m (closest) through 10 m (farthest), including the direction of effects (positive or negative relative to the reference level) and relevant interactions.

## 3.3 Results

### 3.3.1 Japanese experiment

At the closest distance 0.5 m, participants on average moved the text from 0.5 m out to approximately 0.76 m ($b = 0.26$, $p < 0.001$). Average EMM-predicted $FDdeID$ shifts were +0.40 m for EL1 (final $\approx$ 0.90 m), +0.20 m for CL1 ($\approx$ 0.70 m) and +0.19 m for JL1 ($\approx$ 0.69 m), indicating that English-native readers required a farther viewing distance than Chinese- and Japanese-native readers.

At comfortable near distance 2.5 m, participants on average moved the text slightly closer than the initial position, but script complexity effects emerged: simpler glyphs (J1: $b = 0.09$, SD = 0.03, $p < 0.01$; J3: $b = 0.10$, SD = 0.03, $p < 0.01$) were set a bit farther back than complex Kanji. EMM-predicted $FDdeID$ shifts were −0.31 m (final $\approx$ 2.19 m) for EL1, −0.59 m ($\approx$ 1.91 m) for CL1, and −0.70 m ($\approx$ 1.80 m) for JL1, indicating English speakers needed the text slightly farther away (i.e. appearing smaller) to be comfortable, compared to Chinese - and Japanese - native readers.

At comfortable far distance 5 m, overall effects weakened, with only a strong font–complexity interaction reaching significance (J3 under SaB font: $b = 0.33$, $p = 0.02$). EMM-predicted $FDdeID$ shifts under the Black font were +0.27 m (final $\approx$ 5.27 m) for simple characters (J1–J3) versus −1.18 m (final $\approx$ 3.82 m) for complex characters (J4–J6). Under the Thin font, simple characters shifted +0.09 m and complex −1.05 m, yielding a similar $\sim$ 1.14 m viewing - distance gap. Although font weight had only a minor effect, the contrast between simple and complex characters remained the dominant factor.

At extreme far distance 10 m, participants adjusted the text until it was just barely recognizable. Notably, simple - character text (J1–J3) remained legible at farther distances, with $FDdeID$ shifts of approximately +0.30 to +0.56 m. Furthermore, bold fonts amplified this advantage: Katakana J2 in SaB yielded $b = 0.56$ ($p < 0.001$)), and Hiragana J1 in SeB yielded $b = 0.42$ ($p < 0.001$)). In contrast, the thin serif font reduced legibility for simple kana—J2 in SeT required a $b = -0.57$ shift ($p < 0.001$)—underscoring the critical role of stroke weight at very low resolution.

Moreover, EMM - predicted $FDdeID$ values show that under bold font conditions (SaB and SeB), Japanese-native readers tolerated distances beyond 10 m (JL1: +1.64 m, final $\approx$ 11.64 m), whereas Chinese and English readers required the text slightly closer (CL1: −0.14 m, final $\approx$ 9.86 m; EL1: −0.11 m, final $\approx$ 9.89 m). For complex Kanji, JL1 and CL1 converged around 9 m, but EL1 still needed text about 1 m closer. Taken together, these results indicate that familiarity with complex graphs helps maintain their

recognizability via top - down perceptual completion, even when fine stroke details vanish.

In summary (Japanese text), the results from the Japanese VR reading task indicate several key points:

1. First, English speakers needed the text at a slightly greater distance (larger visual size) than Japanese and Chinese readers at the close distances (0.5 m and 2.5 m).

2. Second, simpler Japanese characters (like Kana or simple Kanji) could be read from farther away than complex Kanji (notable at 2.5 m and 5 m), highlighting the impact of visual complexity on legibility distance.

3. Third, at the extreme far distance 10 m, font weight made a crucial difference: heavy strokes significantly extended the legible range for simple characters, whereas thin strokes made recognition much more difficult (requiring closer distances).

4. Finally, Kana (J1 – J2) are native only to Japanese readers, only JL1 participants could identify those at extended distances; by contrast, complex Kanji (J4 – J6) are familiar to both Japanese and Chinese natives (JL1 and CL1), so both groups outperformed EL1 in legibility range.

### 3.3.2 Chinese experiment

In the Chinese text condition the results showed a similar pattern to the Japanese experiment – effects of font weight, character complexity, and some group differences – but with a few notable differences in the role of native-language familiarity.

At 0.5 m, EL1 readers had an average EMM-predicted $FDdeID$ shift of +0.42 m (final $\approx$ 0.92 m), compared to +0.21 m for CL1 and +0.31 m for JL1, showing that English speakers needed nearly twice the viewing distance. A font - weight effect also appeared at 0.5 m: lighter fonts (SeE) produced a small forward shift $b = -0.04$ ($p < 0.01$). Under bold fonts, EL1 readers retreated even more (EL1: +0.47 m; JL1: +0.32 m; CL1: +0.22 m). This means an English speaker confronted with thick, unfamiliar Chinese characters at 0.5 m would move even farther away ( 1.0 m) – possibly the combination of unfamiliar shapes and overwhelming stroke thickness made the close distance uncomfortable, whereas native readers were a bit more at close.

At 2.5 m, very few effects reached significance, indicating that by this comfortable distance, most conditions were similarly manageable. However, when looking at the overall adjusted distances, the familiar pattern remains: EL1 participants still kept the text further out than the others for comfort. The mean predicted positions were approximately –0.38 m (final 2.12 m) for EL1, compared to –0.65 m ( 1.85 m) for JL1 and –0.75 m ( 1.75 m) for CL1. By 5 m, group- and stimulus-level differences re-emerged. The LME showed that JL1 participants placed some texts significantly farther than the grand mean; for C2 (moderate complexity), JL1 placements averaged +0.29 m

($p < 0.001$). Based on the EMM - predicted averages, glyph complexity produced a pronounced bias: across all groups, simple characters (C1) were moved outward by about +0.39 m on average, whereas very complex characters (C4) had to be pulled in by −1.78 m—an overall range difference of over 2 m.

At 10 m, the results echo those of the Japanese experiment in many respects. Overall, participants had to move the text substantially closer (the intercept was $b = -1.47$, $p < 0.001$). Within this general trend, font weight and character complexity played pivotal roles. Mirroring the Japanese experiment, the sans-black font with simple glyphs (SaB × C1) deviated positively by +0.43 m from that mean ($b = 0.43$, $p < 0.01$), meaning these stimuli remained legible at a greater distance. In contrast, Based on the EMM-predicted averages, when glyph complexity peaked (C4), texts of any font had to be moved forward by about 4 m from the initial 10 m to be recognized (C1 = +0.96 m, C4 = −4.1 m).

In summary, the Chinese experiments revealed four key patterns:

1. At the 0.5 m and 2.5 m, native English readers required farther viewing distances than both native Japanese and native Chinese readers, mirroring the findings from the Japanese study.

2. Using the thin font weight on simple glyphs reduced legibility and forced participants to move closer, mirroring the findings from the Japanese study.

3. Simple glyphs remained recognizable at greater distances, whereas complex glyphs necessitated a closer viewing distance.

4. Unlike the Japanese experiment, no clear native-language advantage emerged: native Chinese readers adjusted their viewing distances for complex glyphs almost identically to non-native readers.

### 3.3.3 English experiment

Overall, the effects in English text condition observed here align with the previous experiments, underlining font and familiarity influences, with some nuances since the English alphabet is structurally different from Chinese/Japanese scripts.

At 0.5 m, participants generally moved the text back (LME intercept $b = 0.15$, $p < 0.01$). Thin fonts prompted a small forward adjustment ($b = -0.04$, $p < 0.01$). EMM-predicted $FDdeID$ were −0.01 m for JL1 and +0.05 m for CL1 (both ≈ 0.50 m), whereas EL1 readers shifted +0.41 m (final ≈ 0.91 m), showing that native English readers required substantially more distance than Chinese - and Japanese - native readers to read comfortably at the closest setting.

At 2.5 m, Japanese - native readers made an extra forward adjustment for medium - length English text in Sans - Black (JL1 × SaB × E2: $b = -0.13$, $p < 0.05$). More broadly, bold fonts (SaB, SeB: $b \approx +0.09$–$0.11$, $p < 0.001$) supported a comfortable

viewing distance of $\approx 1.87$ m versus $\approx 1.64$ m for thin fonts (a difference of $\approx 0.23$ m). Finally, EMM-predicted $FDdeID$ values indicate EL1 readers preferred $\approx 2.37$ m, while both CL1 and JL1 readers preferred $\approx 1.46$ m.

At 5 m, the only significant fixed effect was a three-way interaction: JL1 readers in SeB viewing E1 moved the text 0.26 m closer ($b = -0.26$, $p < 0.05$; final $\approx 4.63$ m). Additionally, English-native readers continued to prefer greater distances than CL1 and JL1 across all word lengths. Font weight effects increased: EMM predicted comfortable distances were $\approx 4.77$ m with bold fonts (SaB, SeB) versus $\approx 4.19$ m with thin fonts (SaT, SeT), a 0.58 m gap—indicating that light fonts substantially impaired legibility at 5 m, whereas bold fonts preserved it.

At 10 m, participants showed an average forward shift ($b = -0.84$, $p < 0.01$). The font weight effect was strongly evident: text in bold fonts could be recognized at significantly farther distances than text in thin fonts. The predicted threshold was roughly 9.36 m for bold versus 8.73 m for thin, a difference of $\approx 0.63$ m. This confirms that the trend seen at 2.5 m and 5 m continued: in extreme conditions, light fonts lose clarity much sooner than heavy fonts. We can interpret this as a resolution issue – at 9–10 m in VR, letters subtend very small visual angles, and thin strokes likely fall below a pixel' s width or blend into background, whereas bold strokes still render as visible lines. It is worth noting that all three participant groups ultimately had to bring text closer at 10 m, but consistent with the pattern at other distances, EL1 readers maintained a farther threshold than the others. From the EMM, EL1 could get the text out to 10.56 m for recognition (depending on font), whereas CL1 and JL1 were typically around 7.6–9.3 m for the threshold.

In the English experiment, two key findings emerged:

1. Native English readers consistently preferred—or were able to recognize text from —greater distances than non-native readers (Chinese and Japanese) across all distance conditions (0.5 m–10 m).

2. Font weight was a critical factor: thin fonts invariably required closer viewing, while heavier weights permitted more distant reading. At moderate distances (2.5 m, 5 m), using a bold versus thin font could change the comfortable distance by a few tens of centimeters, and at the extreme distance (10 m) it made over a half-meter difference in legible range.

### 3.3.4 Result Summary

To aid the reader, Figure 3.3 summarizes the main empirical patterns:

- **Language familiarity (10 m).** Native-language text was readable at farther distances than foreign-language text in the Japanese and English experiments, show in Figure 3.3a.

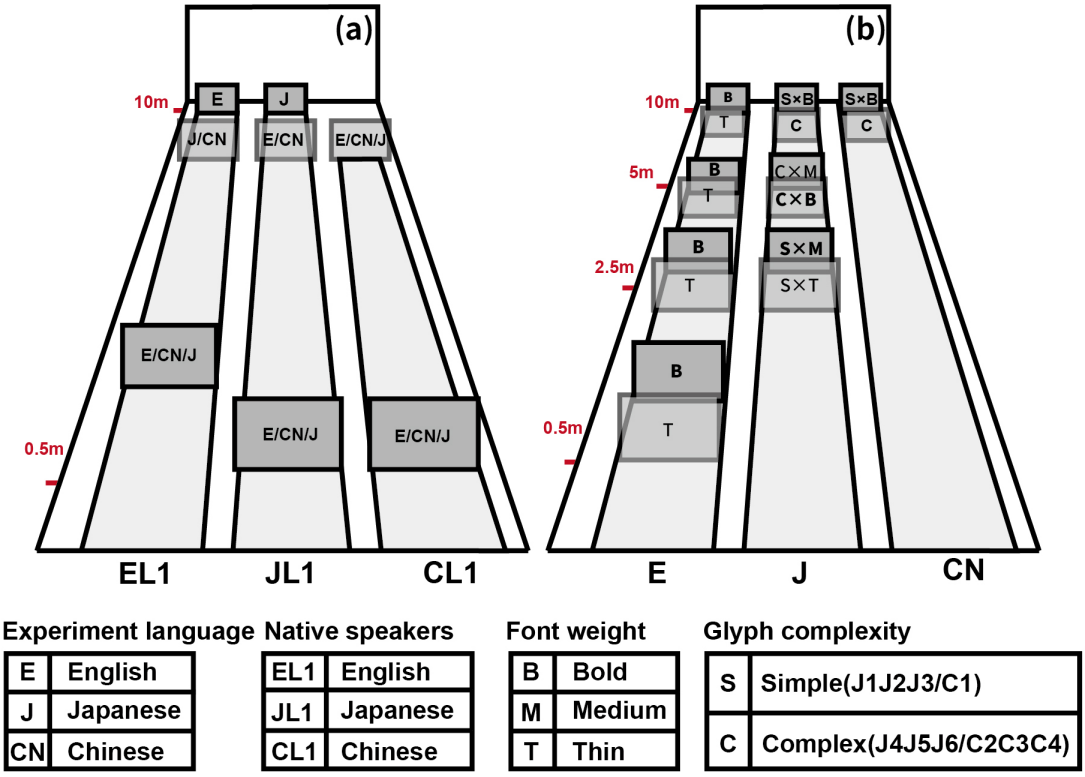



Fig. 3.3: Schematic summary of observed viewing-distance effects in VR. (a) Language familiarity patterns; (b) effects of glyph complexity and font weight.

- **Near distances (0.5 m–2.5 m).** EL1 readers positioned the text farther than CL1 and JL1, show in Figure 3.3a.
- **Glyph complexity.** Simple glyphs were readable at farther distances than complex, high-stroke glyphs (Chinese/Japanese experiments), show in Figure 3.3b.
- **Font weight vs. family.** Bold weights increased readable distance while light weights required closer viewing; family (serif vs. sans-serif) effects were smaller, show in Figure 3.3b.

## 3.4 Discussion of Key Findings

Although research on L1–L2 text perception and related topics is relatively mature, it remains challenging to compare language effects within a shared experimental environment. The present chapter highlights several robust behavioral phenomena in VR reading, as well as a number of novel findings.

### 3.4.1 Effect of Native Language on Perceptual Distance

As shown in Figure 3.5a and b, at 0.5 m and 2.5 m native English readers preferred farther viewing distances than native Chinese and Japanese readers. Because the English-native participants in Japan came from diverse ethnic backgrounds, small anatomical differences (e.g., interpupillary distance) are likely to play only a minor role. A plausible account is that English reading relies on a comparatively large, right-asymmetric perceptual span ($\approx$3–4 letters left, $\approx$14–15 right) and parafoveal preview that guides saccade targeting ($\approx$7–9 letters), as summarized in classic eye-movement work [110] and recent reviews of dynamic reading [111] (schematized in Fig. 3.4). When the text is very near, a word occupies more visual angle and fewer letters fall inside this foveal+parafoveal window; increasing distance restores a preview "buffer" at word beginnings/ends. In contrast, the perceptual span is smaller in Japanese ($\approx$5–7 characters, script-dependent) [112] and narrower in Chinese ($\approx$1 left, 3–4 right) [113], so the benefit of adding extra "buffer" by moving farther is reduced. Although these measures were not obtained in VR, the present results provide some support for this interpretation. Even though prior work has suggested that language does not fundamentally alter basic spatial cognition [114], such language-specific reading habits may still operate in VR and influence spatial perception.

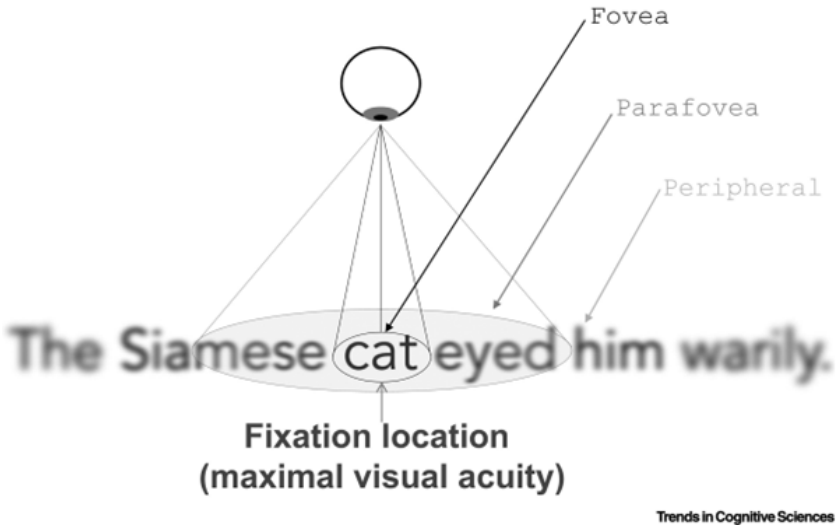



Fig. 3.4: Schematic illustration of foveal and parafoveal processing during reading, highlighting the perceptual span [111].

### 3.4.2 Visual compensation

"Visual compensation" refers to using prior knowledge to complete degraded or missing visual details, a top–down process documented in perception research [115, 116]. In the 10 m condition, as shown in Figure 3.6, native readers of Japanese and English could

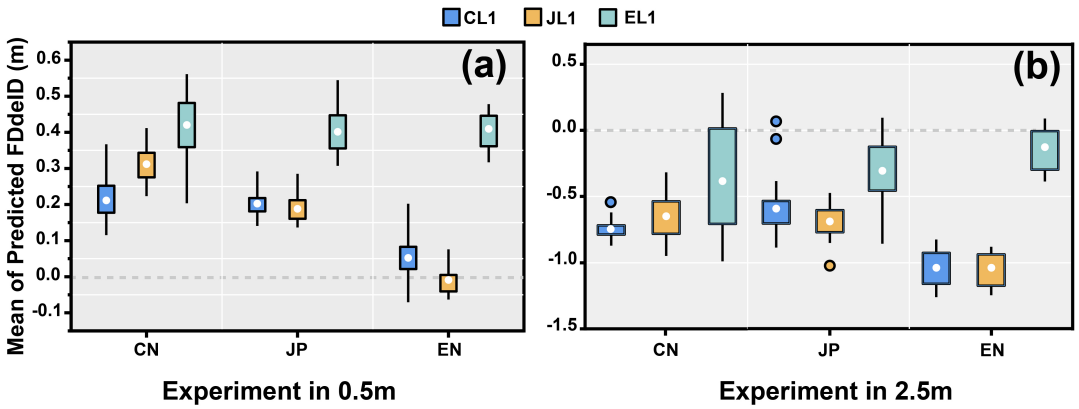



Fig. 3.5: EMM-predicted mean viewing-distance adjustments for native Chinese (CL1), Japanese (JL1), and English (EL1) readers across all language experiments at 0.5 m (panel a) and 2.5 m (panel b).

identify the stimuli at greater distances than non-natives under the same acuity limits. This pattern suggests that familiarity supports top–down completion of blurred strokes and helps preserve legibility when pixel resolution is low.

Interestingly, the same advantage did not appear in the Chinese experiment. Three plausible explanations, all consistent with the data, are considered below. First, the effect is stimulus-dependent: in the Japanese set, compensation is most visible for low-complexity glyphs (J1–J3; Kana and simple Kanji) and weakens as complexity increases (J4–J6; Figure 3.6a). English letters are also low in stroke complexity, which may make compensation more effective (Figure 3.6b). By contrast, the Chinese set consists solely of Hanzi, whose higher within-character stroke density likely requires finer stroke resolution, leaving less room for "filling-in" under blur [71].

Second, the 10 m condition used a subjective visibility/comfort threshold without accuracy constraints. Such a criterion can encourage more conservative choices for dense Hanzi, whereas simpler glyphs (Kana/Latin letters) can still feel legible at longer distances. Notably, in objective identification paradigms, native Chinese readers often show expertise-based advantages consistent with top-down support [116]. This pattern suggests that familiarity-driven completion likely operates in Chinese as well, but the subjective 10,m criterion adopted in this experiment may have attenuated its observable benefit.

Third, differences between Hanzi-only (Chinese) and mixed Kanji–Kana (Japanese) writing may shape long-term strategies: Japanese readers are routinely exposed to strings of low-stroke Kana glyphs (e.g., particles, inflections, furigana), whereas Chinese readers predominantly read continuous Hanzi without an analogous syllabary; extended sequences of extremely low-stroke glyphs are therefore less typical in Chinese text. Although the present design cannot isolate these factors, they suggest concrete hypotheses

for future work.

Overall, the absence of a native-Chinese advantage in 10 m likely reflects an interaction between glyph complexity, stroke density, and the subjective nature of the 10 m task. Future work with larger and more diverse samples, real-word stimuli, and matched Hanzi/Kanji sets is needed to test these mechanisms directly.

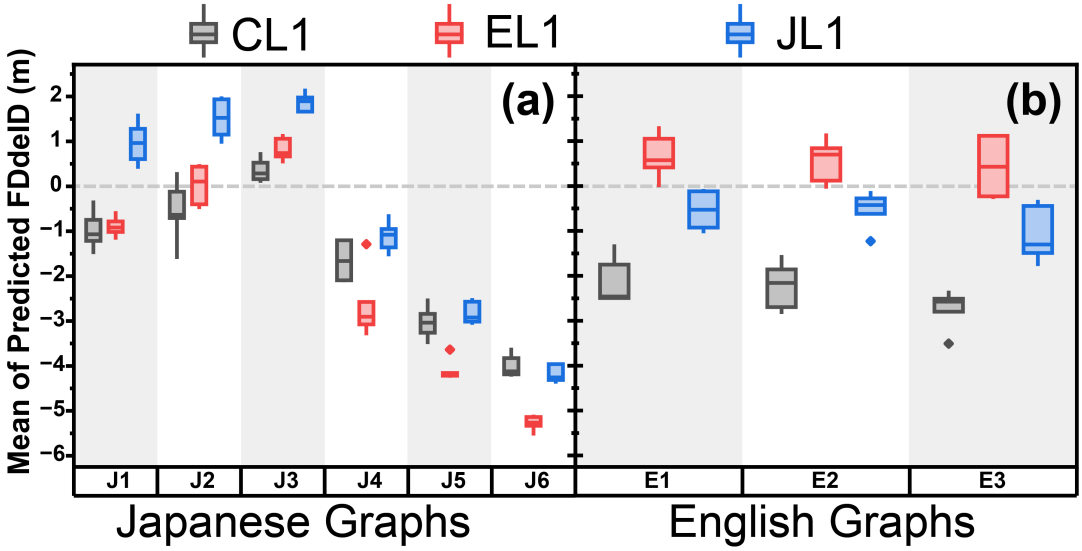



Fig. 3.6: EMM-predicted mean viewing-distance shifts under 10 m for different glyphs and native groups in the Japanese(a) and English(b) experiments.

### 3.4.3 Design Implications for Complex Glyphs

The results indicate that glyph complexity has a profound impact on VR text legibility: characters with more strokes require substantially shorter viewing distances. Under the far (10 m) condition, EMM-predicted $FDdeID$ values in the Chinese and Japanese experiments (Figure 3.7(a) and (b)) reveal a clear divergence: simple glyphs (with fewer strokes) remain recognizable several meters farther out, whereas complex glyphs must be brought markedly closer. Although native readers show a slight visual-compensation advantage, identifying text marginally farther than non-natives, the effect size is small.

In the English experiment (Figure 3.7(c)), comparing short versus long words did not yield a consistent distance advantage, suggesting that word length alone does not replicate the stroke-count effects seen in logographic scripts. Together, these patterns imply that detail-rich text is especially vulnerable to reduced resolution and increased viewing distance. This accords with established visual - science findings: as spatial frequency or display resolution decreases, fine details disappear first. [58, 67] In VR, pixelation and aliasing cause complex strokes to vanish or merge, raising the size threshold for recognition. For designers, these results recommend using simpler, minimally decorated

glyphs—or even abstract symbols—when legibility at a distance is critical.

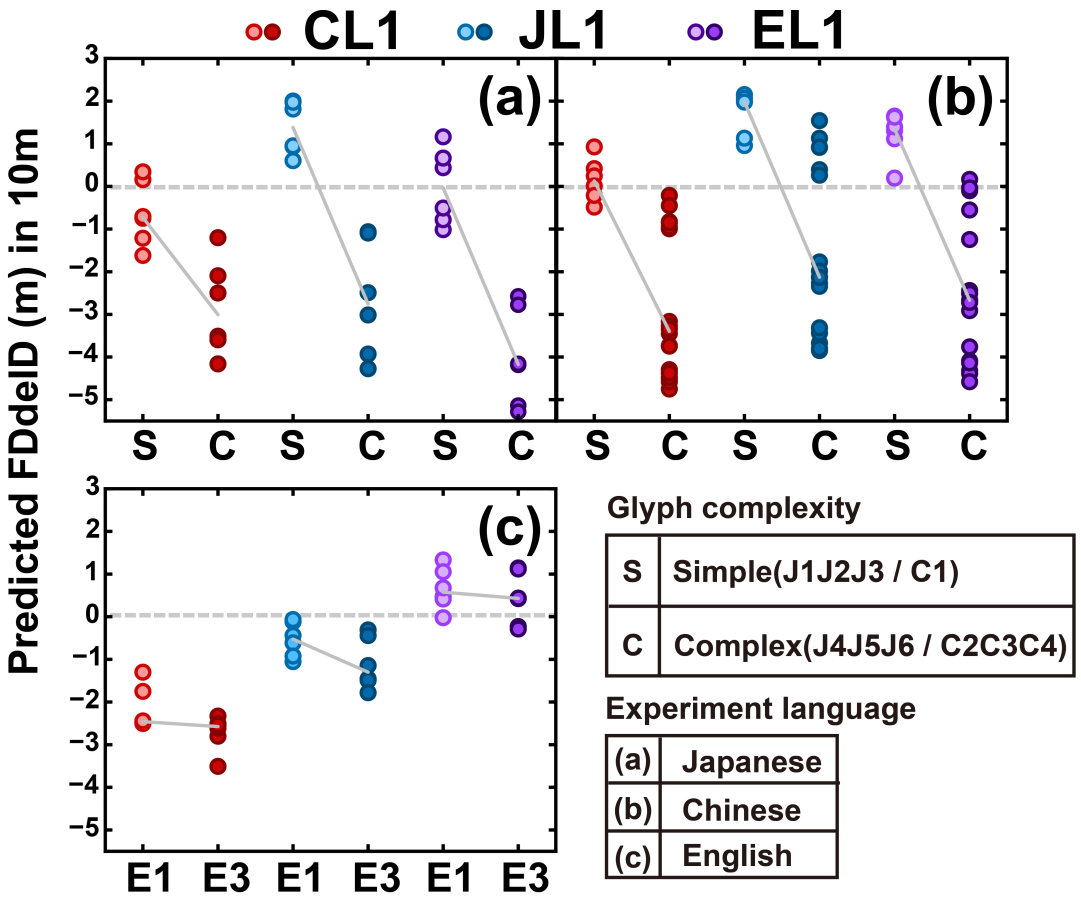



Fig. 3.7: EMM - predicted 10 m viewing - distance shifts across three experiments: simple versus complex glyphs in Japanese (a) and Chinese (b), and short(E1) versus long(E3) words in English (c), for native Chinese, Japanese, and English readers

### 3.4.4 Subtle Effects of Font Weight

Although prior research has demonstrated substantial font effects, EMM predictions from the present models indicate that native language and glyph complexity are the dominant factors driving viewing-distance adjustments, with font weight exerting only conditional influence. Because overall font style had little effect in both experiments, the interaction of weight and complexity was examined by classifying glyphs as simple (J1–J3/C1) or complex (J4–J6/C2–C4), as shown in Figure 3.8.

In the Japanese experiment as shown in Figure 3.8(a–d), simple glyphs in the Thin weight required the shortest distances at 2.5 m (Figure 3.8a), whereas complex glyphs in Bold needed the closer distances at 5 m (Figure 3.8d). However, in the Chinese

experiment as shown in Figure 3.8(e–h), non-native readers clearly needed closer viewing distances for complex glyphs in Figure 3.8(g & h), indicating that heavier weights further impede recognition of intricate Chinese characters for non-native participants.

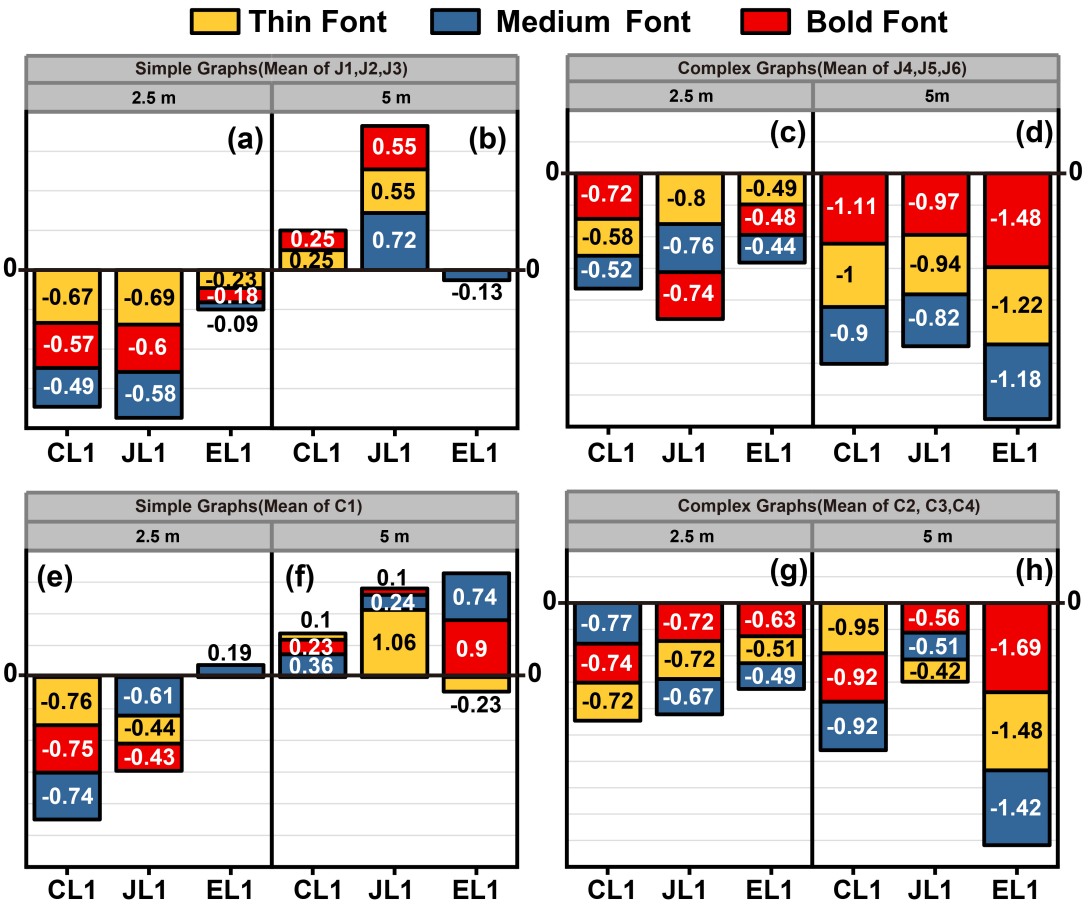



Fig. 3.8: Cumulative EMM-predicted viewing-distance shifts at 2.5 m and 5 m by font weight and native-language group for simple and complex glyphs in the Japanese (a–d) and Chinese (e–h) experiments. Positive values indicate moving the panel farther; negative values indicate moving it closer. For readability, values with $|\Delta| < 0.10$ m are shown as 0 (label omitted).

In the English experiment (Figure 3.9), where glyph complexity is not a factor, six fonts were compared directly. At 2.5 m (Figure 3.9 a & c), Sans - Thin (SaT) demanded the closer distance among non - serif fonts, whereas at 5 m (Figure 3.9 b & d), Serif-Bold (SeB) supported the farthest comfortable viewing distance—results that align with numerous studies [67, 68, 92, 93, 117] showing light weights impair legibility and bold weights enhance readability. Overall, font - weight effects are consistent but modest. This recurring "subtle effect" supports a pixel - threshold hypothesis: in VR' s limited resolution, thin strokes quickly fall below visibility, while bold strokes maintain clear

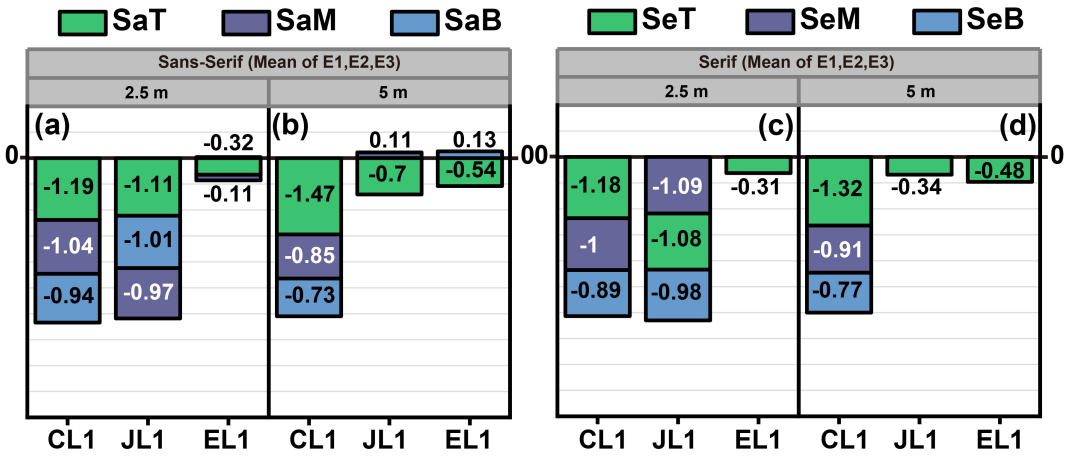



Fig. 3.9: Cumulative EMM-predicted viewing-distance shifts at 2.5 m and 5 m in the English experiment, grouped by font family and native-language group. Positive values indicate moving the panel farther; negative values indicate moving it closer. For readability, values with $|\Delta| < 0.10$ m are shown as 0 (label omitted).

contours, thus extending the legible range by a small but reliable margin.

## 3.5 Chapter Summary

This chapter examined how native language, glyph complexity, and font properties shape readable viewing distance in VR by testing native Chinese, Japanese, and English readers on three scripts across four distances (0.5 m–10 m). Three consistent patterns emerged: (i) language familiarity increased the readable range in the Japanese and English experiments at the far distance (10 m), consistent with top-down compensation, whereas this advantage was not evident with Chinese stimuli; the mechanisms underlying this pattern remain tentative and require further experimental validation; (ii) at near distances (0.5 m–2.5 m), native English readers preferred slightly farther placements, consistent with a larger, right-asymmetric perceptual span; and (iii) simple glyphs and bolder weights remained legible at farther distances, whereas complex characters and light weights required closer viewing, with font-family effects being small.

Building on these findings, several practical guidelines for VR text design can be proposed. For Chinese and Japanese scripts, bolder strokes should be used at greater viewing distances, and complex glyphs should be enlarged to preserve fine details. For English text, light-weight fonts should be avoided, and text size can be modestly increased for non-native readers. For interfaces primarily targeting English-native users, text panels should not be placed too close; adding extra padding around text elements helps maintain a comfortable visual buffer. Together, these guidelines help maintain clear, accessible text across diverse language backgrounds and provide an empirical foundation

for supporting productivity, learning, and social interaction in virtual reality.

# Chapter 4 Understanding Glyph Characteristics on the Legibility for Native and Non-Native Speakers in Virtual Reality

Following the distance-centric analysis in Chapter 3, this chapter examines glyph perception as a primary factor governing text recognizability in VR.

Orthographies differ not only linguistically but also in their visual organization. Hanzi/Kanji compress multiple subcomponents into compact two-dimensional blocks, whereas kana are discrete, block-like syllabic glyphs with minimal internal structure; English, by contrast, distributes information across linear sequences of letters (see Section 2.3.3). Japanese combines logographic Kanji with syllabic kana, whose letterforms are visually discrete yet historically derived from Chinese characters [30]. Such structural contrasts imply different perceptual demands: logographic characters often invite more holistic processing, whereas alphabetic words entail more serial decoding [54, 118]. Meanwhile, although Japanese kana represent syllables, L2 learners may perceive them as visual gestalts similar to Kanji because of their chunked, block-like appearance [53]. These observations highlight the need for a systematic framework to compare cross-linguistic visual-perceptual demands in immersive, digitally mediated environments.

For L2 users, these demands interact with prior experience, yielding processing profiles that diverge from those of native readers [119–121]. Building on this, Chapter 2 established a data-driven complexity scale for Chinese and Japanese characters (Section 2.3.3) and aligned the English materials via word-length tiers. The present chapter introduces a second, orthogonal dimension: glyph similarity. Recognition difficulty depends not only on how much detail a glyph contains (complexity) but also on how easily it can be confused with other glyphs of similar shape. In practice, confusable pairs or sets are assembled within each script: for Chinese and Japanese, characters that share radicals or have near-isomorphic skeletons; for kana, visually similar letterforms; and for English, words with overlapping silhouettes or letter patterns. These sets define low-similarity and high-similarity conditions. Similarity is manipulated while viewing distance (Section 2.3.2) and font factors (style × weight, Section 2.3.4) are held constant or counterbalanced; detailed operational criteria and exemplars are provided in

the Experiment section. This design makes it possible to test how script-internal visual structure and reader background jointly shape glyph recognizability in immersive displays.

## 4.1 Related Work: Cross-Linguistic Cognitive Differences in Native vs. Non-Native Language Processing

### 4.1.1 Cognitive and Behavioral Disparities in L1 vs. L2 Processing

Research comparing native (L1) and second language (L2) processing has evolved significantly since the late 20th century. Early studies emphasized cultural schemata and prosodic awareness as critical factors in L2 comprehension. For instance, L2 readers perform better with culturally familiar content [119], while non-native users adapt their language structures when interacting with AI assistants, prioritizing clarity over fluency [122]. Prosodic processing further differentiates L1 and L2 users: native listeners leverage stress patterns for rapid semantic parsing [120], whereas L2 learners struggle with prosodic cues due to interference from their L1 rhythmic patterns [121, 123].

Advancements in cognitive psychology and educational technology have shifted focus to orthographic differences between writing systems. Studies highlight that L1 and L2 readers exhibit distinct processing strategies rooted in graphemic familiarity, the visual recognition of writing units. For example, Lee *et al.* [124] found that font legibility significantly impacts L2 learners' reading efficiency, while Gauvin *et al.* [125] showed that degraded typography disproportionately hinders non-native speakers. These findings underscore the role of visual ergonomics in L2 acquisition.

Orthographic layout parameters, such as spacing and text directionality, further modulate cross-linguistic processing. While word spacing improves comprehension for L2 readers of Chinese [126], script directionality (e.g., left-to-right vs. right-to-left) shows negligible transfer effects in L2 English reading [127]. Collectively, these studies reveal that L1-L2 disparities arise from both macro-level factors (cultural schemata, prosody) and micro-level visual properties (typography, layout).

### 4.1.2 Glyph Perception in Logographic vs. Alphabetic Systems

Glyphs—the minimal visual units of writing systems—serve as critical interfaces between orthography and cognition. Unlike morphemes or graphemes, glyphs emphasize visual form over linguistic meaning, making them pivotal for investigating cross-linguistic perceptual differences.

Logographic systems (e.g., Chinese Hanzi) compress semantic complexity into spatially nested structures. Despite increased stroke density, complex characters like 贏 (yíng, "to win") show no significant impact on comprehension [128], suggesting holistic visual processing dominates. In contrast, alphabetic systems (e.g., English) exhibit

sequential complexity: longer words like "unbreakable" linearly integrate morphemes, requiring incremental decoding that amplifies cognitive load [64]. This dichotomy reflects fundamental differences in visual information encoding: logographic systems prioritize spatial efficiency, while alphabetic systems demand temporal sequencing [129].

L1 background further modulates glyph perception. Pae *et al.* [130] found that Chinese-native speakers exhibit heightened sensitivity to typographic distortions in English, likely due to their reliance on holistic glyph recognition. Such findings align with Kosaka *et al.* [53], which attributes L2 learners' gestalt-like processing of Japanese Kana to visual chunking strategies. These studies collectively suggest that L1 orthographic experiences shape visual processing biases, with logographic learners favoring global features and alphabetic learners focusing on local details.

### 4.1.3 Synthesis and Research Positioning

Taken together, prior work shows two points clearly. First, display and typography choices in VR influence legibility [28, 49, 56, 131–134]. Second, L1 and L2 readers do not weight visual evidence in the same way, and their performance reacts differently to changes in visual form [124–127, 130].

What is still missing is a direct test, inside VR, of glyph-level factors that are known to matter during recognition, while viewing geometry and font settings are held under control. This chapter therefore concentrates on two orthogonal properties of glyphs: (i) **complexity**, using the tiers defined in Chapter 2, and (ii) **similarity** (confusability among visually close forms within each script). Low- and high-similarity sets are contrasted for Chinese/Japanese characters and for English words, and L1 readers are compared with L2 readers under the same distances and typeface settings. The aim is to isolate how these two properties shape recognition in VR, so that script-aware and reader-aware design guidance can be stated without confounding from viewing distance or font tuning.

Grounded in prior work on L1/L2 reading and orthographic processing, the chapter addresses the following hypotheses across three outcome measures (error rate, invisibility rate, and task completion time):

- **H1 (Native-language advantage).** Native speakers (L1) are expected to show lower error and invisibility rates, as well as shorter task completion times, than non-native readers (L2/LN).
- **H2 (Similarity cost).** High structural similarity between glyphs is expected to increase error and invisibility rates and to prolong task completion time, relative to low-similarity pairs.
- **H3 (Complexity cost).** Higher glyph complexity (stroke count or word length) is expected to increase error and invisibility rates and to prolong task completion time.

- **H4 (Font influence).** Font style and weight are expected to modulate these effects, such that certain typefaces or weights yield better legibility than others.

## 4.2 Experiment Design

### 4.2.1 Virtual Environment

A 14 × 2.5 × 2.5 m virtual corridor environment was created, in which stimuli were presented in three languages (Chinese, Japanese, and English) within a 570 × 200 mm virtual text box. Based on previous research [135], the background of the text box was set to black with 50% transparency, and the text color was white. The text box displayed either two characters (Chinese/Japanese) or two words (English). A prompt box positioned below the user' s camera view randomly indicated one of the characters or words shown in the text box. Participants used the trigger buttons on the left and right Quest Touch Pro controllers to indicate whether the prompted character or word appeared on the left or right side of the text box. The scene was illuminated with directional lighting to ensure adequate visibility throughout. A schematic of the scene is shown in Figure 4.1.

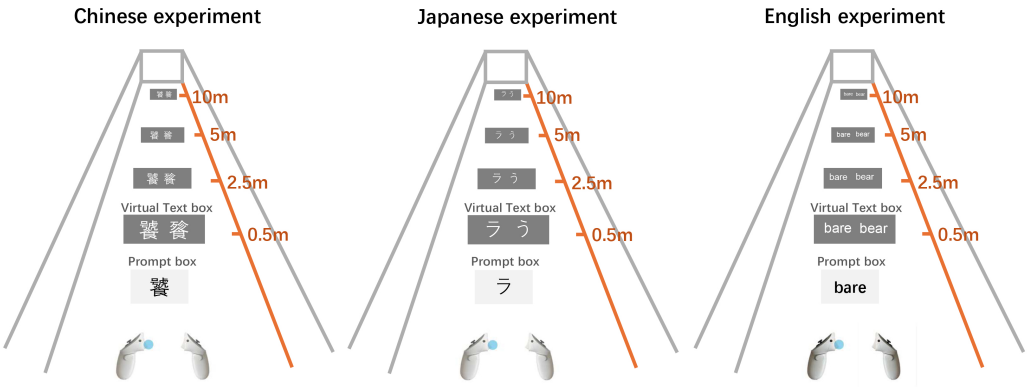



Fig. 4.1: Schematic of the experiment scene. This figure illustrates the virtual environment in which participants viewed a text box containing either two Chinese/Japanese characters or two English words at varying distances (0.5 m, 2.5 m, 5 m, and 10 m).

### 4.2.2 Independent variable

A between-participant design was used. The four independent variables were as follows:

1. Glyph Complexity and Similarity (11 groups for Japanese, 9 groups for Chinese,

and 7 groups for English)

2. Fonts (6 types)
3. Distance (4 distances)
4. Native language (3 types)

### Glyph

**Glyph Complexity** The complexity tiers established in Chapter 2 (Section 2.3.3) are adopted here without rederiving the measure. For Chinese and Japanese characters (Hanzi/Kanji), visual complexity is indexed by a composite of stroke count and pixel footprint; for English, visual demand is aligned via word-length tiers. This shared scaling provides a script-agnostic baseline to which the new similarity manipulation introduced in this chapter can be added. Within each complexity tier (Section 2.3.3), items are further divided into similar and dissimilar groups according to script-specific visual rules. The intention is to combine readers' intuitive sense of visual similarity with reproducible classification criteria.

**Glyph Similarity** Within each complexity tier, items are further divided into similar and dissimilar groups using script-specific visual rules. The goal is to align readers' intuitive sense of "look-alike" forms with reproducible criteria.

For Chinese characters (Chinese Hanzi/Japanese Kanji), glyphs are block-like and vary in both structural layout and stroke count. Accordingly, structural configuration is combined with a stroke-count threshold:

1. **Structure.** Similar pairs share the same layout (e.g., left–right, top–bottom, enclosure); dissimilar pairs have different layouts.
2. **Stroke count.** Similar pairs differ by $\leq 2$ strokes; dissimilar pairs differ by $> 2$ strokes.

These rules were first applied automatically to the Hanzi/Kanji database and then verified by manual inspection. Example: 明 vs. 朋 (same left–right layout; stroke difference $\leq 2$) are similar; 明 vs. 品 (different layout; stroke difference $> 2$) are dissimilar.

Japanese kana glyphs are visually simple (most have no more than 3–4 strokes) and highly stylized; curves dominate hiragana, whereas straight segments dominate katakana. To avoid confounding by style, similarity is defined in terms of canonical stroke sequence (order and dominant direction): two kana are treated as similar if their canonical stroke sequences match apart from minor stylistic variation; otherwise, they are treated as dissimilar. Example: あ and お share a comparable horizontal → vertical → hook sequence and are similar; あ and し differ in sequence and are dissimilar.

For English words, visual complexity rises primarily with length and affixation. Words are stratified by length—short ($\leq$ 3), medium (4–8), long ($\geq$ 9)—and, within each stratum, label pairs as similar if any of the following holds:

- Anagrams differing only in letter order (e.g., bare vs. bear);
- Same base with a single-letter prefix/suffix added or removed (e.g., reference vs. preference);
- Same core morpheme with one internal segment altered (e.g., implementation vs. improvisation).

Pairs failing these criteria are labeled dissimilar.

Therefore, crossing the complexity tiers from Chapter 2 with the similarity rules above yields the glyph conditions used in this chapter. To ensure diversity and reduce fatigue, five non-overlapping glyph sets were randomly sampled and presented within each classification group. The number of groups per language is:

- Japanese: 11 groups (complexity $\times$ similarity)
- Chinese: 9 groups (complexity $\times$ similarity)
- English: 7 groups (complexity $\times$ similarity)

Table 4.1 enumerates the character sets by language, complexity tier, and similarity label.

### Distance

As in Chapter 2, four viewing distances were defined within the virtual environment: 0.5 m, 2.5 m, 5 m, and 10 m. These distances correspond, respectively, to the minimum distance at which characters remain recognizable, a comfortable reading distance for nearsighted individuals, a suitable reading distance for farsighted individuals, and the maximum distance at which character outlines are barely distinguishable.

### Font

Given the vast number of available fonts and the substantial stylistic differences between languages, this experiment adopted the same Google Noto superfamily [72–77] described in Chapter 2. These fonts were selected for their consistent styling across languages and their broad adoption in practical applications.

### Native language

For the native language, following the recruitment of the native language as described in Chapter 2, this experiment also used 10 native Chinese speakers, 10 native English

Table. 4.1: Character Glyph Classification by Feature and Similarity Across Chinese, Japanese, and English

| **Language** | **Feature** | **Character sets** | **Similarity** |
|---|---|---|---|
| Japanese | Kana | ぬね　びひ　ララ　ソン　ババ | Similar （J1） |
| | | おす　フナ　とト　ゆン　ラガ | Dissimilar （J2） |
| | Kanji($C \leq 0.5$) | 人入　八入　丁了　士土　力刀 | Similar （J3） |
| | | 山小　二又　三女　大上　才十 | Dissimilar （J4） |
| | Kanji($0.5 < C \leq 2$) | 究完　査直　没沿　板牧　固囲 | Similar （J5） |
| | | 呉茶　呂皇　果直　徒逆　造挑 | Dissimilar （J6） |
| | Kanji($2 < C \leq 4$) | 説設　関開　親頻　載戦　寧箋 | Similar （J7） |
| | | 格唇　渇夏　瘍棄　徳質　塊圏 | Dissimilar （J8） |
| | Kanji($C > 4$) | 議譲　麗麓　闇簡　謄騰　鶴鶏 | Similar （J9） |
| | | 瞳襲　騒麗　露璽　欄覇　観覆 | Dissimilar （J10） |
| | Mixed | 人お　矢ソ　質三　襲ナ　護技 | Dissimilar （J11） |
| Chinese | Hanzi($C \leq 0.5$) | 丁了　刀刁　入人　厂广 | Similar （C1） |
| | | 乃了　一乙　入卜　工之 | Dissimilar （C2） |
| | Hanzi($0.5 < C \leq 2$) | 朵朵　伶冷　匂勾　氷永 | Similar （C3） |
| | | 汽冉　沁由　冰決　忾汴 | Dissimilar （C4） |
| | Hanzi($2 < C \leq 4$) | 萎萋　設説　董堇　間問 | Similar （C5） |
| | | 鵐奠　鴻葛　齊疸　感庸 | Dissimilar （C6） |
| | Hanzi($C > 4$) | 麤龘　穱穱　饕餮　魑魎 | Similar （C7） |
| | | 賽関　孽镰　鵚騙　麋譨 | Dissimilar （C8） |
| | Mixed | 丁魑　乃觧　了鸿　魑尤 | Dissimilar （C9） |
| English | 0–3 letters | no on　me he　cat cap | Similar (E1) |
| | | on or　run ran | |
| | | go he　on me　to up | Dissimilar (E2) |
| | | cat dog　sky big | |
| | 4–8 letters | bare bear　seam seem　script scrip | Similar (E3) |
| | | bread break　dream dram | |
| | | duce bear　green sight　script crept | Dissimilar (E4) |
| | | animal beauty　gallery kitchen | |
| | over 8 letters | reference preference　demonstration determination | Similar (E5) |
| | | illustration imagination　consideration conversation | |
| | | implementation improvisation | |
| | | rehabilitation disappointment　communication accomplishment | Dissimilar (E6) |
| | | encouragement documentation　conference magnificent | |
| | | appreciation conference | |
| | Mixed | for reference　on magnificent | Dissimilar (E7) |
| | | cap determination　sky accomplishment | |
| | | to preference | |

speakers, and 10 native Japanese speakers. The participants were also assigned to the native language and non-native language projects. Therefore, the final participants in this Chinese experiment (10 native Chinese speakers, 5 native Japanese speakers, 5 native English speakers), the participants in the Japanese experiment (10 native Japanese speakers, 5 native Chinese speakers, 5 native English speakers), and the participants in the English experiment (10 native English speakers, 5 native Chinese speakers, 5 native Japanese speakers)

### 4.2.3 Task

As shown in Figure 4.1, The task required participants to identify the character or word displayed in the prompt box and determine its position (left or right) within the text box on the glyph canvas. For example, in the Chinese experiment, the character " 饢 " appears in the prompt box and is positioned on the left side of the canvas displaying " 饢餮 ", prompting a left trigger press. Similarly, in the Japanese and the English experiments, " ラ " and "bare" are positioned on the left, necessitating the same selection . This setup enables the assessment of language-specific glyph recognition across diverse linguistic backgrounds.

### 4.2.4 Procedure

Participants were welcomed, briefed on the purpose and procedure, and gave informed consent in accordance with the Institutional Review Board of Hokkaido University. As compensation, each participant received a QUO card valued at JPY 1,000 upon completing the experiment. They then completed VR training and a pilot session until proficient with the HMD and Quest Touch Pro controllers.

The experiment was conducted in Japanese, Chinese, and English. Each participant performed one comprehension task in their L1 and one in a randomly assigned non-native language (L2). For each language, 10 native and 10 non-native participants (5 from each of the other two groups) were split into forward (0.5 m→10 m) or reverse (10 m→0.5 m) distance orders.

In the main task, participants completed one trial for each combination of 4 distances (0.5, 2.5, 5, 10 m) × 6 fonts across the per-language glyph-group conditions (JP = 11, CN = 9, EN = 7). A single block lasted approximately 10–15 min to avoid fatigue or cybersickness (symptoms tend to increase after about 10 min of VR exposure) [136], with a 10 s rest every 50 trials. Glyph pairs (similar/dissimilar) were displayed at the set distances, and participants selected the matching glyph using the controller triggers; they used the grip button to indicate visual difficulty. After a 2–3 min break, they repeated the task in the second non-native language. Total experiment time was $\leq 30$ min.

Following HMD removal, participants provided subjective feedback. Figure 4.2 illustrates the overall flow and task steps. All behavioral measurements, participant demo-

graphics, and representative recordings of the experimental procedure are available in a supporting data repository on Open Science Framework (OSF [137]).

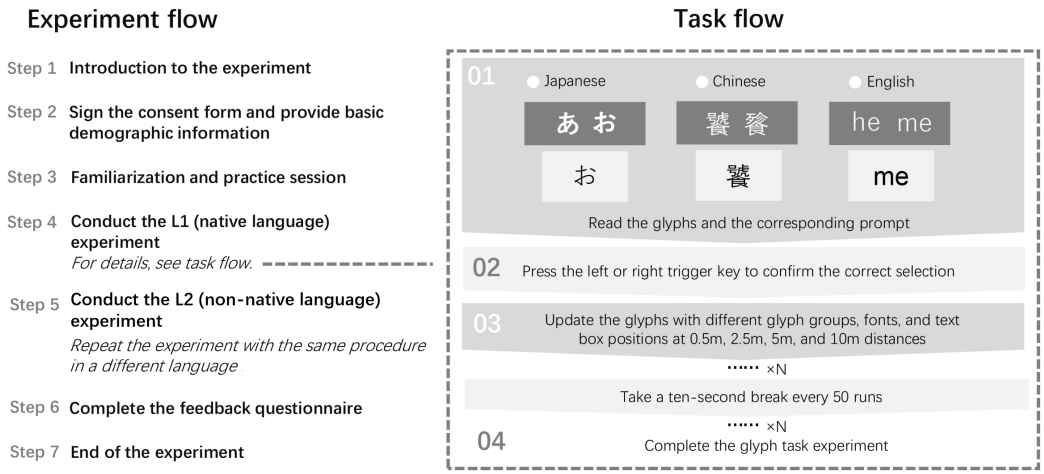



Fig. 4.2: Experiment Flow and Task Flow Diagram. This figure illustrates the overall flow of the experiment, starting with the introduction and consent signing, followed by practice sessions, the main tasks in L1 (native language) and L2 (non-native language), and concluding with participant feedback. The task flow section on the right details the specific steps participants followed during each trial, including reading the glyph prompt, selecting the matching glyph using the Quest Touch Pro controllers, and taking breaks between sets of trials.

### 4.2.5 Data Collection and Analysis

#### Data Collection

Accuracy was assessed by comparing each selection (left/right) with the glyph shown and the prompt. Incorrect choices defined the *Error Rate*, whereas grip-button presses on entirely illegible glyphs defined the *Invisibility Rate*. Selection latency was recorded as *Task Completion Time*.

Participants used the Quest Touch Pro triggers to select the matching glyph, guessing when the target was unclear but still visible and pressing the grip button when it was unreadable. For every trial, the system logged glyph distance, font, response type (left, right, invisible), and completion time.

In total, 12,960 valid trials were recorded:

1. Japanese: 11 glyph groups $\times$ 4 distances $\times$ 6 fonts $\times$ 20 participants = 5,280
2. Chinese: 9 glyph groups $\times$ 4 distances $\times$ 6 fonts $\times$ 20 participants = 4,320

3. English: 7 glyph groups × 4 distances × 6 fonts × 20 participants = 3,360

This dataset provides a comprehensive view of participants' performance and visual perception across conditions.

**Data Analysis**

For statistical analysis, three outcome measures were defined: error rate, invisibility rate, and task completion time. Because the task was simple and both error and invisibility events were rare, standard regression models would have yielded unstable estimates. Accordingly, analyses were conducted within SPSS's Generalized Linear Model (GLM) framework (ANOVA within the generalized linear model class) to evaluate between-group differences, and post-hoc pairwise comparisons were performed using both Tukey's HSD and estimated marginal means (EMMs), with effects of negligible $\eta^2$ excluded from interpretation. Significant interaction effects were further examined using error-bar plots and interaction graphs to visualize key trends and to mitigate challenges posed by sparse error data.

Task completion time served as a supplementary outcome measure. The GLM framework was again applied to assess group effects, followed by Tukey's HSD and EMM-based post-hoc analyses, omitting effects with negligible $\eta^2$. For significant two-way interactions, observed trends were visualized; for more complex three-way interactions, linear mixed-effects (LME) models were fitted. By integrating GLM results, LME analyses, and graphical visualizations, the analyses explored potential reasons why certain language groups performed better or worse under specific interaction conditions. Statistical significance was set at $p < .05$. All GLM and LME outputs, as well as supporting analysis scripts, are available in a supporting repository on OSF [137].

## 4.3 Result

To minimize the impact of individual differences and keep the task straightforward, the experimental design was simplified so that even L2 participants with no prior knowledge of the target language could succeed through careful visual observation. All incorrect and invisible responses were inspected, and task completion times were recorded to ensure data reliability and to rule out accidental button presses. Because correct responses predominated, error and invisibility rates were very low, which in turn violated the homogeneity-of-variance assumption for a standard GLM-based ANOVA. These rare events, however, are highly informative for legibility. Accordingly, estimated marginal means (EMMs; pairwise comparisons) and post-hoc multiple-comparison tests were used, complemented by visualizations and practical interpretation, to provide a more comprehensive and nuanced analysis.

The study included four viewing distances, each corresponding to a distinct sensory experience. Treating distance merely as a single independent factor in a pooled analysis

would therefore have limited interpretive value. Instead, and in order to focus more closely on glyph characteristics, between-group effects were examined separately across languages and distance levels. The results are organized and discussed by language and viewing distance in the sections that follow.

### 4.3.1 Error Rates

While the experiment included a large number of trials overall, the distribution of trials across multiple conditions (e.g., Glyph, Font, and Native) resulted in relatively few trials for each specific combination of variables. Combined with the low overall error rates, this can be considered a low-event-rate design rather than a traditional small-sample experiment. The total number of errors under different distances and conditions is summarized in Table 4.2. In the table 4.2, "Errors" represents the total number of errors for each distance, while the "Glyph" column lists values for each language in the order of C1–C9, J1–J11, and E1–E7. To enhance readability, similar glyphs are highlighted in bold (C1, C3, C5, and C7 for Chinese; J1, J3, J5, J7, and J9 for Japanese; and E1, E3, and E5 for English). The "Font" column lists values for fonts in the order of SaB, SaM, ST, SeB, SeM, and SeE/SeT (SeE for Japanese and Chinese, and SeT for English). The "Native" column lists values for native speakers in the order of Chinese, Japanese, and English participants, while "Total" represents the total number of trials for each distance.

Table. 4.2: Error Distribution Across Different Distances, Glyphs, Fonts, and Native Languages

| **Language** | **Distance** | **Errors** | **Glyph** | **Font** | **Native** | **Total** |
|---|---|---|---|---|---|---|
| CN | 0.5 m | 14 | **1**,1,**6**,0,**5**,0,**0**,1,0 | 0,0,5,1,4,4 | 7,3,4 | 918 |
| | 2.5 m | 16 | **1**,1,**0**,4,**3**,3,**2**,1,1 | 2,3,2,2,6,1 | 7,4,5 | 918 |
| | 5 m | 26 | **3**,0,**5**,4,**8**,0,**6**,0,0 | 2,8,5,3,2,6 | 10,9,7 | 918 |
| | 10 m | 85 | **2**,0,**17**,8,**25**,5,**23**,4,1 | 13,16,19,13,15,9 | 46,20,19 | 918 |
| JP | 0.5 m | 17 | **2**,0,**4**,1,**3**,0,**0**,0,**3**,3,1 | 5,3,1,4,3,1 | 6,4,3 | 1122 |
| | 2.5 m | 12 | **1**,0,**1**,0,**1**,0,**1**,0,**5**,0,1 | 1,3,3,2,1,2 | 7,5,0 | 1122 |
| | 5 m | 22 | **4**,0,**5**,0,**5**,0,**2**,0,**4**,1,1 | 4,7,6,2,2,1 | 11,7,4 | 1122 |
| | 10 m | 89 | **11**,0,**4**,0,**16**,1,**16**,4,**21**,16,0 | 19,9,11,19,13,18 | 22,33,34 | 1122 |
| EN | 0.5 m | 15 | **1**,1,**7**,0,**5**,0 | 3,2,4,2,3,1 | 4,6,5 | 714 |
| | 2.5 m | 5 | **0**,1,**2**,0,**1**,1 | 1,1,1,1,0,1 | 0,3,2 | 714 |
| | 5 m | 10 | **2**,0,**5**,0,**3**,0 | 1,0,2,1,3,3 | 5,5,0 | 714 |
| | 10 m | 22 | **4**,1,**6**,3,**4**,3 | 1,3,3,5,3,7 | 12,8,2 | 714 |

From the table 4.2, it can be observed that error rates for Chinese and Japanese primarily occur under the 10 m condition. In contrast, English experiments exhibit lower overall error rates, with slightly higher rates under 0.5 m and 10 m.

**English Experiment**

At the farthest distance (10 m), only a main effect of Native reached significance [$F(2, 714) = 8.50$, $p < .01$, $\eta^2 = 0.023$). At the intermediate distance (5 m), significant main effects of Native, $F(2, 714) = 5.312, p < .01, \eta^2 = 0.015$, and Glyph, $F(6, 714) = 4.703, p < .01, \eta^2 = 0.038$, were found, along with significant Native × Glyph, $F(12, 714) = 2.214, p < .05, \eta^2 = 0.036$, and Font × Glyph, $F(30, 714) = 1.575, p < .05, \eta^2 = 0.062$, interactions. ANOVA results were examined separately for each viewing distance. Under the closest distances (0.5 m, 2.5 m), no main or interaction effects reached significance (all $p > .05$), indicating uniformly low error rates when stimuli were highly visible.

Under 5 m and 10 m, both Post-hoc contrasts confirmed that native English speakers(EL1) made significantly fewer errors than both Chinese speakers(CL1) and Japanese speakers(JL1) ($MD < 0, p < .05$).MD means Mean Difference, calculated as the difference in error rates between groups.Since errors were coded as "1" and correct responses as "0", Comparing the former and the latter, $MD > 0$ indicates a higher error rate in the former, and $MD < 0$ indicates a lower error rate.

Interaction effects (Glyph × Native and Font × Glyph) produced more nuanced patterns under 5 m. Estimated marginal means showed elevated error rates for visually similar glyph sets (E1: $p = .012$, $\eta^2 = .012$; E3: $p < .001$, $\eta^2 = .027$; E5: $p = .017$, $\eta^2 = .011$). Specifically:

- For E1, JL1 had higher error rates than both CL1 and EL1 (JL1 vs. CL1/EL1: $MD = 0.07/0.07, p < .05$).
- For E3, JL1 and CL1 had higher error rates than EL1 ($MD = 0.07/0.10, p < .05$).
- For E5, CL1 had higher error rates than EL1 ($MD = 0.07, p < .05$).

These patterns suggest that logographic - language natives are particularly prone to errors when encountering similar English glyphs. Moreover, at 5 m, Serif Medium (SeM) and Serif Thin (SeT) fonts further increased error rates for E3 glyphs relative to other glyphs (all $MD > 0, p < .05$).

In summary, at close viewing distances (0.5 m, 2.5 m), error rates remain uniformly low regardless of glyph or font. As viewing distance increases (5 m, 10 m), however, a native - language effect becomes apparent, with English natives making fewer errors than Chinese and Japanese natives. This gap is especially pronounced at 5 m, where the combination of visually similar glyphs and Serifthin font details disproportionately disrupts participants from logographic backgrounds (CL1 and JL1). These patterns suggest that natives of logographic languages may process English text in a manner more akin to symbolic scripts.

**Japanese Experiment**

ANOVA results were examined separately by viewing distance. Under the farthest distance (10 m), there were significant main and interaction effects for Glyph [$F(10, 1122) = 10.495, p < .01, \eta^2 = 0.086$], Glyph × Native [$F(20, 1122) = 1.902, p < .05, \eta^2 = 0.033$], and Glyph × Font [$F(50, 1122) = 1.637, p < .01, \eta^2 = 0.068$]. At the intermediate distance (5 m), Glyph [$F(10, 1122) = 2.674, p < .01, \eta^2 = 0.023$] and Glyph × Native [$F(20, 1122) = 1.777, p < .05, \eta^2 = 0.031$] remained significant. Under 2.5 m, only Glyph reached significance [$F(10, 1122) = 2.047, p < .05, \eta^2 = 0.018$], and at the closest distance (0.5 m), both Glyph [$F(10, 1122) = 2.302, p < .01, \eta^2 = 0.020$] and Glyph × Native [$F(20, 1122) = 1.718, p < .05, \eta^2 = 0.030$] were significant.

Overall, glyph identity was the primary driver of error rates. Specifically, visually similar glyphs consistently produced higher error rates than non-similar ones, with this effect most pronounced at greater distances. For example, under 10 m the non-similar glyphs J4 and J6 yielded lower error rates than the similar sets J5, J7, and J9 ($MD < 0, p < .05$). As viewing distance decreased, the similarity effect attenuated: although the most complex similar glyph J9 still exceeded the complex non-similar J10 in error rate ($MD > 0, p < .05$), other simpler or moderately complex glyphs no longer differed significantly.

Glyph complexity played a secondary role. At 10 m, the simpler glyph J3 evoked fewer errors than the medium-complexity similar J5 ($MD < 0, p < .05$), whereas the simple non-similar kana J2 did not differ from more complex glyphs J4, J6, or J8 ($p > .05$). Thus, similarity rather than stroke count primarily determined error susceptibility.

Interactions with Native language and Font added further nuance under specific conditions. At far distances (5 m/10 m), native English speakers were particularly challenged by complex glyphs: for J9 ($\eta^2 = 0.016$) and J10 ($\eta^2 = 0.011$), EL1 error rates significantly exceeded those of CL1 and JL1 (J9: $MD = 0.22/0.20$; J10: $MD = 0.17/0.20$, both $p < .01$). Conversely, native Chinese participants showed higher errors on simple similar glyphs at 5 m—in the J3 comparison ($\eta^2 = 0.019$), CL1 errors were greater than both EL1 and JL1 ($MD = 0.12/0.13, p < .001$). Moreover, bold serif styling (SeB) amplified similarity effects at 10 m: for J5 ($p < .001, \eta^2 = 0.028$), error rates in SeB surpassed all fonts except SaB ($MD > 0, p < .05$). Even at closer distances (2.5 m/0.5 m), English native speakers remained more error-prone on the most complex glyph J9 (0.5 m: $\eta^2 = 0.016; EL1 vs. JL1/CL1 : MD = 0.10/0.10, p < .01$).

**Chinese Experiment**

ANOVA results were examined for each viewing distance. Under the farthest distance (10 m), Glyph was the sole significant factor influencing error rates [$F(8, 918) = 10.131, p < .01, \eta^2 = 0.081$]. At the intermediate distance (5 m), Glyph again reached significance [$F(8, 918) = 3.568, p < .01, \eta^2 = 0.030$]. No effects emerged at the nearer distance (2.5 m; all $p > .05$). At the closest distance (0.5 m), both Glyph [$F(8, 918) =$

$2.927, p < .01, \eta^2 = 0.025$] and the Glyph × Native interaction [$F(16, 918) = 1.718, p < .05, \eta^2 = 0.032$] were significant.

Overall, glyph identity drove error susceptibility. At 10 m, among glyphs of equivalent stroke count, visually similar forms (e.g., C5, C7) elicited higher error rates than non-similar counterparts of the same complexity (C4, C6, C8; $MD > 0, p < .01$). When complexity differed, more intricate shapes (C3, C5, C7) produced more errors than simple forms (C1, C2; $MD < 0, p < .01$), and medium-complexity similar glyphs (C3, C5) exceeded the error rate of complex but non-similar glyphs (C8, C9; $MD > 0, p < .05$).

At the comfortable viewing distance 5 m, similarity remained the dominant influence: the medium-complex similar glyph C5 showed elevated errors relative to non-similar glyphs C2, C6, C8 ($MD > 0, p < .05$). Even at the closest distance 0.5 m, similar shapes continued to challenge participants (e.g., C3 vs. C4/C6/C7/C9; $MD > 0, p < .05$). Notably, at 0.5 m the Glyph × Native interaction revealed that on the complex similar glyph C5($\eta^2 = 0.018$), native Chinese speakers(CL1) made significantly more errors than both native Japanese(JL1) and native English speakers(EL1) (both $MD = 0.08, p < .001$), suggesting a nuanced interplay between script familiarity and visual similarity.

### Interaction Analysis

The main factors affecting error rates were Native and Glyph and their interaction. Trends across distances are shown in Figure 4.3, with parts (A)–(C) depicting Glyph × Native interactions for Chinese, Japanese, and English experiments.

From these plots:

- **Chinese:** As shown in Figure 4.3(A), Although Glyph × Native was not significant ($p > .05$), Chinese and Japanese speakers had similar errors on C3, C5, and C7, while English speakers erred more on C3 but less on C5 and C7.
- **Japanese:** Figure 4.3(B) shows that similar glyphs (J1, J3, J5, J7, J9) yielded high errors, especially J9. Native Japanese speakers had the lowest errors; English speakers' errors rose with complexity; Chinese speakers' errors remained steady.
- **English:** As illustrated in Figure 4.3(C), overall errors were lower, with peaks at E1, E3, and E5 (E3 highest). Non-native English speakers showed greater sensitivity to similarity.

Similar glyphs were consistently more error-prone, especially for Chinese and Japanese speakers, while complexity affected native English speakers most in the Japanese experiment. Lower errors for complex glyphs in the Chinese experiment may reflect marking them as "invisible" rather than guessing.

Significant Glyph–Font interactions ($p < .05$) appeared in the English (5 m) and Japanese (10 m) experiments, despite non-significant Font main effects ($p > .05$). Figure

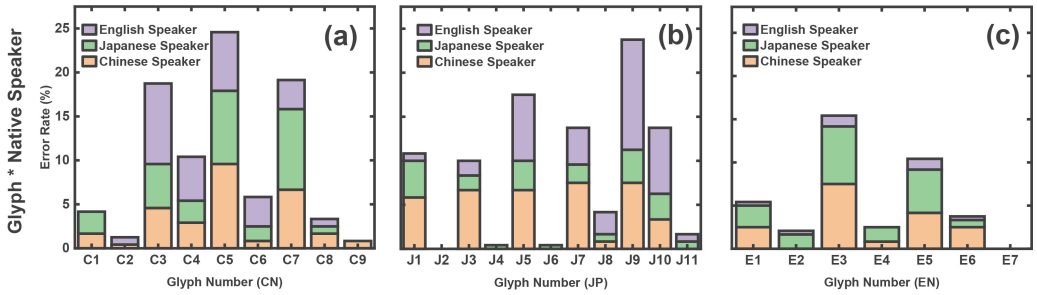



Fig. 4.3: Error rates for Chinese, Japanese, and English Experiments.Parts (A), (B), and (C) depict the interaction between Glyphs and Native language groups. Glyphs C1–C9, J1–J11, and E1–E7 represent increasing complexity, with odd-numbered glyphs (e.g., C3, J5, E3) being similar and even-numbered glyphs (e.g., J2, J4) dissimilar. In the Chinese experiment (A), glyphs C3, C5, and C7 show the highest error rates, with English speakers performing worse on C3 but better on C5 and C7 compared to Chinese and Japanese speakers. In the Japanese experiment (B), J9 exhibits the highest error rates, with native Japanese speakers showing consistently lower errors than others. Chinese speakers show balanced errors, while English speakers' errors increase with complexity. In the English experiment (C), overall error rates are lower, but E1, E3, and E5 show peaks, especially for Chinese and Japanese speakers, with E3 being the most error-prone glyph.

4.4 shows error bars: both experiments show higher errors for similar glyphs, but the Japanese experiment had larger font-related fluctuations. In English (5 m), SeM and SeT slightly increased E3 errors; in Japanese (10 m), SeB/SeE peaked at J1, SaB/SeB at J5, SaT/SeT at J7, and SaB/SeB at J9, suggesting serif fonts affect hiragana and font weight matters for complex glyphs.

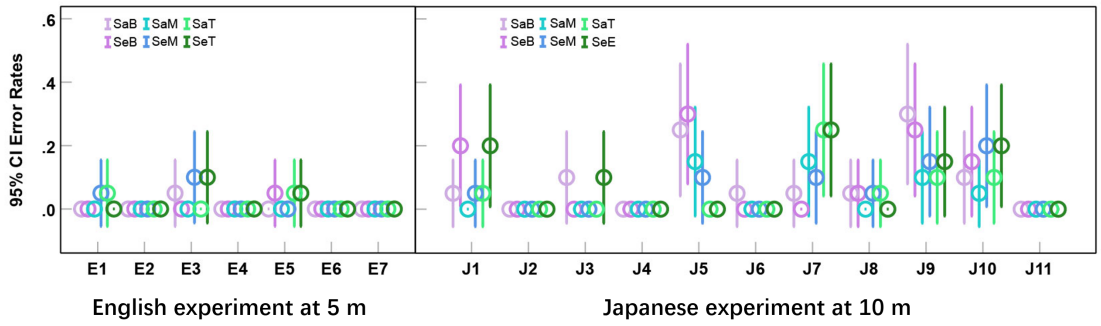



Fig. 4.4: Error bar plots (95% confidence intervals) for Glyph-Font interactions in the English experiment (5 m) and Japanese experiment (10 m).

### Summary of Error Rates

In summary, the error-rate findings with respect to the hypotheses are as follows:

**Hypothesis 1: Native speakers exhibit lower error rates compared to non-native speakers.** Supported in the English and Japanese experiments. In the Chinese experiment, native and non-native speakers showed comparable overall error rates and both groups exhibited stable errors on similar glyphs, indicating that glyph similarity affects Chinese participants regardless of native status.

**Hypothesis 2: Similar glyphs increase error rates.** Confirmed across all languages, with peaks at C5 (Chinese), J9 (Japanese), and E3 (English). Non-native English speakers (CL1 and JL1) had higher error rates on similar English glyphs than native speakers, suggesting that logographic - script users rely on visual similarity in alphabetic scripts in a comparable manner.

**Hypothesis 3: Glyph complexity affects error rates.** Partially supported and less influential than similarity. Elevated errors occurred at C4 (Chinese) and J10 (Japanese). Native English speakers' errors rose with complexity in the Japanese experiment but fell in the Chinese experiment, possibly because highly complex Chinese glyphs ($C > 4$) at the farthest distance were marked as "invisible" rather than guessed.

**Hypothesis 4: Font styles affects error rates.** Partially supported. Significant Glyph $\times$ Font interactions appeared in English (5 m) and Japanese (10 m), yet glyph similarity remained the dominant factor.

### 4.3.2 Invisibility Rates

The overall invisibility rates across the Chinese, Japanese, and English experiments is summarized in Figure 4.5. A one-way ANOVA revealed a significant main effect of language ($F(2, 12957) = 80.677, p < .001, \eta^2 = 0.012$). Given the low occurrence of invisible responses, the effect size ($\eta^2$) is small; however, the actual influence of language may still be meaningful.

Post-hoc comparisons showed that the Chinese group exhibited a significantly higher invisibility rates than both the Japanese group ($MD = 0.02, p < .001$) and the English group ($MD = 0.05, p < .001$). Additionally, the Japanese group had a significantly higher invisibility rates than the English group ($MD = 0.02, p < .001$).Similar to error rates, invisibility responses were coded as "1" (invisible) and "0" (visible). Therefore, $MD > 0$ indicates a higher invisibility rate in the first group, while $MD < 0$ indicates a lower invisibility rate.

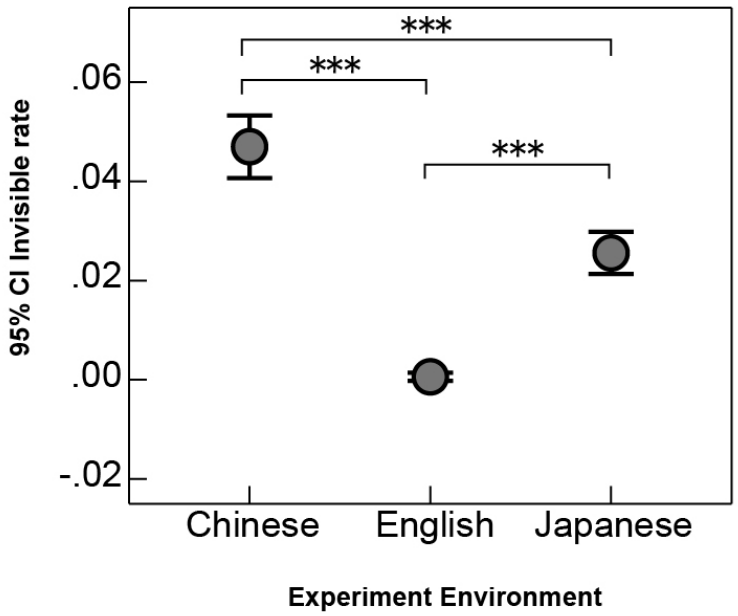



Fig. 4.5: Overall invisibility rates Across Different Language Experiment Environments. The x-axis represents different language experiment environments, while the y-axis indicates the 95% confidence interval of the mean invisibility rates. Overall, the Chinese group had the highest invisibility rates, followed by the Japanese group, while the English group' s invisibility rates were nearly zero.

In addition, the total number of invisible instances across distances was compiled for each language, together with counts under various conditions (e.g., glyphs, fonts, and native languages), as shown in Table 4.3. Similarly, and consistent with the structure of Table 4.2, similar glyphs (e.g., C1, C3, C5 for Chinese; J1, J3, J5 for Japanese; E1, E3,

E5 for English) are highlighted in bold for easier readability.

Table. 4.3: Invisible Times Distribution Across Different Distances, Glyphs, Fonts, and Native Languages

| Language | Distance | Invisibles | Glyph | Font | Native | Total |
|---|---|---|---|---|---|---|
| CN | 0.5 m | 0 | **0**,0,**0**,0,**0**,0,**0**,0,0 | 0,0,0,0,0,0 | 0,0,0 | 918 |
| | 2.5 m | 5 | **0**,0,**0**,1,**1**,0,**2**,1,0 | 0,2,2,1,0,0 | 3,0,2 | 918 |
| | 5 m | 7 | **0**,0,**0**,0,**1**,0,**4**,2,0 | 2,2,1,1,0,1 | 0,0,7 | 918 |
| | 10 m | 191 | **4**,0,**5**,0,**40**,21,**58**,63,0 | 37,33,30,30,29,32 | 79,47,65 | 918 |
| JP | 0.5 m | 0 | **0**,0,**0**,0,**0**,0,**0**,0,**0**,0,0 | 0,0,0,0,0,0 | 0,0,0 | 1122 |
| | 2.5 m | 0 | **0**,0,**0**,0,**0**,0,**0**,0,**0**,0,0 | 0,0,0,0,0,0 | 0,0,0 | 1122 |
| | 5 m | 3 | **0**,0,**0**,0,**0**,0,**1**,0,**2**,0,0 | 0,0,1,0,1,1 | 0,0,3 | 1122 |
| | 10 m | 132 | **4**,0,**2**,1,**4**,5,**20**,7,**49**,43,1 | 22,27,20,18,21,24 | 51,45,36 | 1122 |
| EN | 0.5 m | 0 | **0**,0,**0**,0,**0**,0,0 | 0,0,0,0,0,0 | 0,0,0 | 714 |
| | 2.5 m | 0 | **0**,0,**0**,0,**0**,0,0 | 0,0,0,0,0,0 | 0,0,0 | 714 |
| | 5 m | 0 | **0**,0,**0**,0,**0**,0,0 | 0,0,0,0,0,0 | 0,0,0 | 714 |
| | 10 m | 2 | **0**,0,**1**,0,**0**,0,1 | 0,0,01,0,1 | 0,2,0 | 714 |

The invisibility rates in the Chinese and Japanese experiments were primarily concentrated under the 10 m condition, where the number of invisible instances increased dramatically. In contrast, English experiments showed an invisibility rates of less than 0.001, likely due to accidental inputs, meaning that invisibility was virtually non-existent. Therefore, English data was excluded from further discussion of invisibility rates. From a practical perspective, invisibility under 0.5 m and 2.5 m conditions in the Chinese experiment is not expected to occur, rendering these data unreliable and excluded. Under the 5 m condition, the data show 7 invisible instances in Chinese and 3 in Japanese. In the Japanese experiment, invisible instances occurred for Japanese native speakers under glyph J7 with the SaT font, and for two English native speakers under glyph J9 with the SeE and SaM fonts. In the Chinese experiment, invisible instances were associated with glyphs C5, C7, and C8, predominantly for English native speakers, and mainly with Sans fonts. Among these glyphs, J7, J9, C5, and C7 are categorized as similar glyphs, while C8 is considered a complex glyph. This indicates that even under the 5 m (comfortable distance) condition, similar glyphs, particularly those with higher complexity, may affect visibility.

### Chinese Experiment

Since invisibility events were concentrated primarily under the 10 m condition, between-group effects were examined at this distance. The analysis revealed a large main effect of Glyph, $F(8, 918) = 53.574, p < .01, \eta^2 = 0.318$, and a smaller but significant Glyph $\times$ Native interaction, $F(16, 918) = 2.26, p < .01, \eta^2 = 0.038$.

Post-hoc comparisons indicated that Glyph complexity determined invisibility: more complex glyph were less clearly visible. Specifically, simple similar glyphs (C1, C3) had lower invisibility rates than complex similar glyphs (C5, C7; $MD < 0, p < .001$),

and likewise, simple non-similar glyphs (C2, C4) were clearer than complex non-similar glyphs (C6, C8; $MD < 0, p < .001$). Although C5 was less visible than the equally complex non-similar C6 ($MD = 0.16, p < .01$), the absence of a difference between C7 and C8 indicates that overall complexity, rather than similarity, is the primary driver of invisibility.

Moreover, the Glyph × Native interaction reached significance for some conditions. For C6 ($\eta^2 = 0.021$), English speakers had higher invisibility than Chinese and Japanese speakers (MD = 0.32/0.28, $p < .01$). For C7 ($\eta^2 = 0.015$), English speakers again showed higher invisibility than Chinese speakers (MD = 0.27, $p < .001$). For C8 ($\eta^2 = 0.012$), Japanese speakers exceeded Chinese speakers in invisibility (MD = 0.23, $p < .01$).

Despite some variability, native speakers consistently showed lower invisibility rates, suggesting either greater reliance on informed guessing under uncertainty or sharper visual discrimination of glyph forms compared with non-native readers.

### Japanese Experiment

A between-group ANOVA showed that Glyph [$F(10, 1122) = 40.851, p < .01, \eta^2 = .267$] and the Glyph × Native interaction [$F(20, 1122) = 3.096, p < .01, \eta^2 = .052$] were the primary drivers of invisibility rates.

Glyph complexity was the dominant driver of invisibility rates: the simpler glyphs J1–J6 were far more visible than the highly complex J7, J9, and J10 ($MD < 0, p < .05$). Within the highest-complexity tier, visual similarity added to difficulty—J7 was less visible than J8 ($MD = 0.14, p < .01$)—whereas J9 and J10 did not differ.

The Glyph × Native interaction emerged for the most challenging forms. At J7 ($\eta^2 = 0.023$), Chinese speakers had higher invisibility rates than Japanese speakers ($CL1 vs JL1 : MD = 0.30, p < .001$). At J9 ($\eta^2 = 0.040$), both Chinese and English speakers exhibited higher invisibility than Japanese speakers (CL1 vs JL1: $MD = 0.40, p < .01$; EL1 vs JL1: $MD = 0.18, p < .01$). The fact that Japanese natives show lower invisibility rates hints that they may prefer inferring ambiguous glyphs over labeling them invisible, or that their perceptual sensitivity to glyph details is stronger than that of non-native speakers.

### Interaction Effects

Overall, Glyph and its interaction with native language were the main factors influencing invisibility rates. To further examine these relationships, stacked area graphs were created to compare the effects of Glyph and native language, as shown in Figure 4.6.

The visualization results revealed comparable patterns between the Chinese and Japanese experiments regarding glyphs and native languages. In the Chinese experiment, glyphs C5, C6, C7, and C9 exhibited higher invisibility rates, while in the Japanese experiment, glyphs J7, J9, and J10 showed higher invisibility rates.

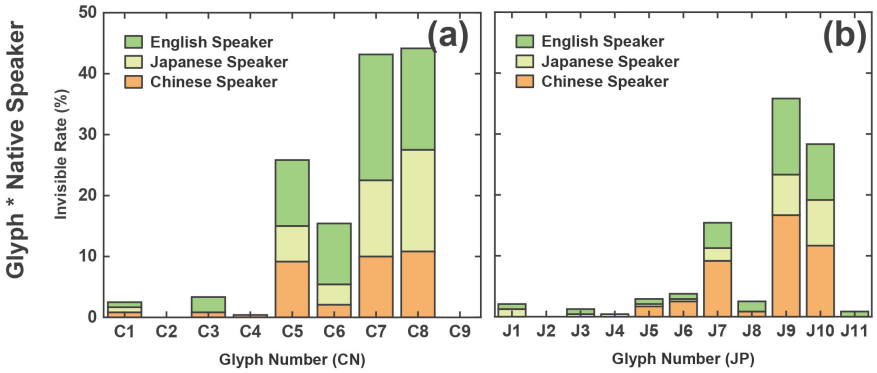



Fig. 4.6: Invisibility Rates in Chinese and Japanese Experiments. Part (A) illustrates the invisibility rates for glyphs C1 to C9 in the Chinese experiment, grouped by native language (Chinese, Japanese, and English speakers).Part (B) depicts the invisibility rates for glyphs J1 to J11 in the Japanese experiment, similarly grouped by native language. Glyphs C5, C6, C7, and C9 showed higher invisibility rates in the Chinese experiment, while glyphs J7, J9, and J10 exhibited higher invisibility rates in the Japanese experiment.

Further analysis indicated differences across native language groups. In the Chinese experiment, Japanese native speakers had lower invisibility rates for glyph C5 compared to Chinese and English native speakers, while English native speakers had higher invisibility rates for glyph C6. Glyphs C7 and C8 displayed relatively consistent invisibility rates across all native groups. In the Japanese experiment, Japanese native speakers consistently exhibited the lowest invisibility rates. However, for glyphs J7, J9, and J10, Chinese native speakers demonstrated higher invisibility rates compared to Japanese native speakers, slightly exceeding those of English native speakers.

In summary, glyph complexity is the primary factor influencing invisibility rates, although glyph similarity also plays a role. In the Chinese experiment, English native speakers had slightly higher invisibility rates compared to Chinese and Japanese native speakers. Conversely, in the Japanese experiment, Japanese native speakers had lower invisibility rates than both Chinese and English native speakers.

### Summary of Invisibility Rates

Based on the hypotheses, the following conclusions are drawn for invisibility rates: **Hypothesis 1: Native speakers exhibit lower invisibility rates compared to non-native speakers.**This hypothesis was supported. In the Chinese experiment, English native speakers exhibited higher invisibility rates compared to Chinese and Japanese native speakers. Similarly, in the Japanese experiment, despite the use of numerous Chinese glyphs, Chinese native speakers had higher invisibility rates compared to Japanese native

speakers. These findings confirm that native speakers generally exhibit lower invisibility rates than non-native speakers.

**Hypothesis 2: Glyph similarity is a major factor influencing invisibility rates.**This hypothesis was partially supported. While similar and complex glyphs, such as J7 and J9, showed some influence on invisibility rates in the Japanese experiment, glyph similarity was not the dominant factor. For instance, J8, which shares the same complexity level as J7, exhibited a lower invisibility rates, suggesting that glyph similarity has a minor but not predominant effect.

**Hypothesis 3: Glyph complexity is a major factor influencing invisibility rates.**This hypothesis was confirmed. Glyph complexity was found to be the primary factor affecting invisibility rates in both Chinese and Japanese experiments. Complex glyphs such as C3, C4, C7, and C8 in the Chinese experiment, and J7, J9, and J10 in the Japanese experiment, consistently exhibited higher invisibility rates compared to simpler glyphs like C1, C2, J1, and J2.

**Hypothesis 4: Fonts influence invisibility rates.**This hypothesis was not supported. Results showed no significant impact of fonts or their interactions on invisibility rates in either the Chinese or Japanese experiments.

In conclusion, glyph complexity is the primary factor influencing invisibility rates for both Chinese and Japanese speakers, whereas native speakers exhibited lower invisibility rates than non-native speakers, particularly for complex glyphs (e.g., C7 and C8, J9 and J10).

### 4.3.3 Task Completion Time

Prior to the experiment, participants were instructed to "select as quickly as possible". During each trial, mean task completion time was recorded. Table 4.4 summarizes mean task completion time across viewing distances, glyphs, fonts, and native-language groups. MT denotes the mean completion time at each viewing distance, and Total denotes the grand mean across 0.5 m–10 m; all values are reported in deciseconds (1 ds = 0.1 s) for ease of comparison. The ordering and structure of the conditions follow those in Tables 4.2 and 4.3. As in Table 4.2, similar-looking glyphs are shown in bold to facilitate readability.

For inferential analysis, a full-factorial GLM ANOVA was first applied at each viewing distance to assess main and interaction effects, followed by post-hoc comparisons on significant main effects to establish initial trends. Next, those two-way interactions that reached significance were visualized. Finally, guided by the GLM main and two-way interaction results, linear mixed-effects (LME) regression was employed to explore why certain language groups performed differently under specific three-way interaction conditions. As with the GLM, separate LME models were run for each experiment (Chinese, Japanese, English) and for each distance (0.5 m–10 m). All data, analyses, and code are available on the Open Science Framework (OSF) [137] for full transparency.

Table. 4.4: Mean Task Completion Time Across Different Distances, Glyphs, Fonts, and Native Languages

| Language | Distance | MT (ds) | Glyph(ds) | Font(ds) | Native(ds) | Total(ds) |
|---|---|---|---|---|---|---|
| CN | 0.5 m | 14 | **12**,10,**15**,13,**16**,15,**18**,16,11 | 15,14,13,15,15,14 | 12,13,20 | 17 |
| | 2.5 m | 15 | **12**,11,**16**,13,**19**,14,**18**,17,12 | 15,14,15,16,14,14 | 13,13,21 | 17 |
| | 5 m | 17 | **13**,11,**18**,14,**25**,17,**26**,19,12 | 18,16,16,18,17,18 | 14,16,25 | 17 |
| | 10 m | 23 | **18**,13,**28**,20,**29**,27,**29**,28,14 | 22,22,24,23,23,24 | 20,23,29 | 17 |
| JP | 0.5 m | 13 | **14**,10,**14**,11,**14**,12,**14**,12,**18**,14,11 | 13,13,13,15,13,13 | 11,12,17 | 16 |
| | 2.5 m | 14 | **13**,10,**15**,11,**15**,12,**16**,13,**21**,14,11 | 13,14,14,14,13,13 | 11,12,19 | 16 |
| | 5 m | 15 | **15**,11,**15**,11,**15**,13,**17**,14,**25**,17,11 | 15,14,15,16,15,16 | 12,13,21 | 16 |
| | 10 m | 22 | **19**,13,**18**,13,**26**,23,**29**,23,**33**,34,15 | 22,22,22,22,22,23 | 17,22,28 | 16 |
| EN | 0.5 m | 16 | **14**,12,**17**,15,**21**,16,14 | 16,16,16,16,16,15 | 14,14,17 | 16 |
| | 2.5 m | 14 | **13**,12,**16**,13,**19**,15,13 | 14,15,15,14,14,14 | 14,13,16 | 16 |
| | 5 m | 15 | **13**,12,**17**,14,**19**,16,12 | 16,15,15,14,15,15 | 15,14,1.6 | 16 |
| | 10 m | 18 | **17**,14,**23**,17,**25**,18,14 | 17,17,19,19,17,10 | 16,18,19 | 16 |

Although reaction time is not treated as direct evidence for the primary conclusions—because participant fatigue from repeated trials may have influenced response speed—it offers valuable supplementary insight. In particular, reaction time helps to interpret the trends observed in error and invisibility rates and provides additional support for the main findings.

### GLM Model Analysis

**English Experiment** ANOVA results by viewing distance showed that at the farthest distance (10 m), Glyph [$F(6, 714) = 31.802, p < .01, \eta^2 = 0.211$], Font [$F(6, 714) = 3.257, p < .01, \eta^2 = 0.022$], and Native [$F(6, 714) = 8.267, p < .01, \eta^2 = 0.023$] were all significant. At the intermediate distance (5 m), only Glyph [$F(6, 714) = 23.988, p < .01, \eta^2 = 0.168$] and Native [$F(6, 714) = 7.036, p < .01, \eta^2 = 0.019$] reached significance. Under 2.5 m, Glyph remained the sole significant factor [$F(6, 714) = 21.294, p < .01, \eta^2 = 0.152$]. At the closest distance (0.5 m), both Glyph [$F(6, 714) = 24.520, p < .01, \eta^2 = 0.063$] and Native [$F(6, 714) = 24.161, p < .01, \eta^2 = 0.171$] were significant. No interaction terms attained significance in any condition.

Across 10 m, 5 m, and 2.5 m, visually similar glyphs consistently required more processing time than dissimilar ones, particularly the longest glyph E5. Under 10 m, E3 and E5 were slower than E1, E2, E4, E6, and E7 ($MD > 0, p < .05$). Because completion time is measured in seconds, $MD > 0$ denotes slower performance (longer times), whereas $MD < 0$ denotes faster performance (shorter times). At 5 m, E3, E5, and E6 slower than E1 and E2 ($MD > 0, p < .05$), with E4 also completing slower than E2 ($MD > 0, p < .05$). Under 2.5 m, E5 remained the slowest across all comparisons ($MD > 0, p < .05$). This pattern indicates that similarity prolongs recognition time at any distance, while as distance decreases, string length exerts an increasingly strong effect—longer words occupy a larger visual angle and thus take longer to process.

The font effect at 10 m was small but detectable: SeT required more time than SaB,

SaM, and SeM ($MD > 0, p < .05$). Interestingly, native English speakers were slower than Chinese and Japanese speakers at 0.5 m (EN vs. CN/JP: $MD > 0, p < .01$), and also at 10 m (EN vs. CN: $MD = 0.27, p < .001$) and 5 m (EN vs. JP: $MD = 0.18, p < .001$), contrary to H1, which predicted that native speakers would respond fastest.

**Japanese Experiment** Separate ANOVAs by viewing distance showed that at 10 m, Glyph [$F(10, 1122) = 36.338, p < .01, \eta^2 = .245$], Native [$F(2, 1122) = 8.267, p < .01, \eta^2 = .102$], and their interaction [$F(20, 1122) = 3.890, p < .01, \eta^2 = .065$] were all significant. At 5 m, the same three effects remained strong (Glyph: $F(10, 1122) = 28.953, \eta^2 = .205$; Native: $F(2, 1122) = 117.856, \eta^2 = .174$; Glyph × Native: $F(20, 1122) = 7.450, \eta^2 = .117$). Under 2.5 m, Glyph ($F(10, 1122) = 17.222, \eta^2 = .133$), Native ($F(2, 1122) = 99.136, \eta^2 = .150$) and their interaction ($F(20, 1122) = 3.567, \eta^2 = .060$) were again significant. Even at the closest distance (0.5 m), Glyph ($F(10, 1122) = 15.496, \eta^2 = .121$), Native ($F(2, 1122) = 104.924, \eta^2 = .158$), Font ($F(5, 1122) = 3.819, \eta^2 = .017$) and Glyph × Native ($F(20, 1122) = 2.914, \eta^2 = .049$) all exerted reliable effects.

Across all distances, native Japanese and Chinese speakers outperformed English speakers: at 10 m, CL1 was faster than JL1 and EL1 ($MD < 0, p < .01$), and at 5 m–0.5 m both CL1 and JL1 remained faster than EL1 ($MD < 0, p < .01$). Glyph type showed two consistent patterns. At the greatest distance (10 m), processing time rose with complexity: the simple dissimilar Kana J2 was completed more quickly than the dissimilar moderate/complex glyphs J6, J8, and J10 ($MD < 0, p < .01$). At closer distances (5 m, 2.5 m, 0.5 m), similarity rather than complexity dominated: within each same complexity tier,for example, J1 vs. J2, J3 vs. J4, J9 vs. J10 at 5 m; J9 vs. J10 at 2.5 m; and J1 vs. J2, J3 vs. J4, J5 vs. J6, J9 vs. J10 at 0.5 m, similar glyphs required significantly more time ($MD > 0, p < .05$). Complexity played a secondary role, emerging only for the most intricate glyphs: the similar most - complex J9 lagged behind its simpler peers like J1, J3, J5 and J7 ($MD > 0, p < .05$), whereas the dissimilar most - complex J10 took longer only compared to dissimilar simple forms like J2 or J4($MD > 0, p < .05$).

The Glyph × Native interaction further highlighted that, as glyphs grew more complex (J6–J10), English speakers took longer than both Japanese and Chinese speakers (EL1 vs. CL1/JL1: $MD > 0, p < .05$). This pattern held even at nearer distances for complex or similar forms (e.g., J5–J10 at 5 m; J3, J5, J7–J10 at 2.5 m; and J3, J5, J7–J10 at 0.5 m), indicating that non-logographic readers consistently struggled more with intricate or confusable glyphs. In sum, at the farthest distance CL1 showed the fastest performance, but as text became clearer, EL1 required longer processing times—especially for complex or similar glyphs—contrary to the expectation in H1 that native speakers would respond fastest.

**Chinese Experiment** Separate ANOVAs by viewing distance revealed that at 10 m, Glyph [$F(8, 918) = 35.3, p < .001, \eta^2 = .235$], Native [$F(2, 918) = 47.114, p < .001, \eta^2 =$

.093] and their interaction [$F(16, 918) = 1.774, p < .05, \eta^2 = .030$] were all significant. At 5 m, Glyph [$F(8, 918) = 61.547, p < .01, \eta^2 = .349$], Native [$F(2, 918) = 156.631, p < .01, \eta^2 = .254$], Font [$F(5, 918) = 3.078, p < .01, \eta^2 = .016$] and their interaction [$F(16, 918) = 8.385, p < .01, \eta^2 = .128$] remained strong. Under 2.5 m, all three—Glyph [$F(8, 918) = 26.241, \eta^2 = .186$], Native [$F(2, 918) = 139.124, \eta^2 = .233$] and Glyph × Native [$F(16, 918) = 6.105, \eta^2 = .096$]—continued to influence performance (all $p < .01$), with Font also reaching significance [$F(5, 918) = 3.631, \eta^2 = .019$]. Even at the closest distance (0.5 m), Glyph [$F(8, 918) = 20.086, \eta^2 = .149$], Native [$F(2, 918) = 163.257, \eta^2 = .262$] and their interaction [$F(16, 918) = 3.998, \eta^2 = .065$] were reliable predictors ($p < .01$).

Overall, Glyph, Native, and their interaction remained significant with moderate–large effect sizes. At the farthest distance (10 m), response time scaled with complexity—more intricate glyphs took longer—while similarity made no additional difference within the same complexity tier. At the intermediate distance (5 m), complexity still dominated, but visually similar pairs (C3 vs. C4, C5 vs. C6, C7 vs. C8) introduced additional delay. At nearer distances (2.5 m and 0.5 m), only complexity slowed recognition; similarity no longer added time.

In addition, a consistent native-language advantage emerged: participants whose native script matched the stimuli (Chinese and Japanese speakers) recognized glyphs more rapidly than those without such familiarity. The Glyph × Native interaction mirrored this ordering for nearly every glyph (except the simplest C1 and C2), indicating that logographic familiarity conferred a speed advantage across character types and viewing conditions.

### Interaction Analysis

Guided by the ANOVA results, mean task completion times were computed for each combination of categorical factors. To illustrate the interaction effects, the relationship between Glyph and Native language was visualized in Figure 4.7. Panels (A), (B), and (C) present the Glyph × Native language interaction for the three experiments.

The figures reveal that, across all languages, similar glyphs generally require more time for recognition. In the Japanese experiment, complex glyphs, particularly intricate Kanji characters, demand significantly longer recognition times. This effect is moderately noticeable in Chinese and Japanese, while in English, the impact is minimal.

Figure 4.7(A–C) shows that native English speakers (EL1) require more time to complete tasks in the Chinese and Japanese experiments than native Japanese (JL1) and Chinese (CL1) speakers. In the Japanese experiment, EL1' s completion time increases as Kanji complexity rises. In the English experiment, all three native groups (Chinese, Japanese, and English) take slightly longer to recognize similar glyphs, though the overall differences remain small.

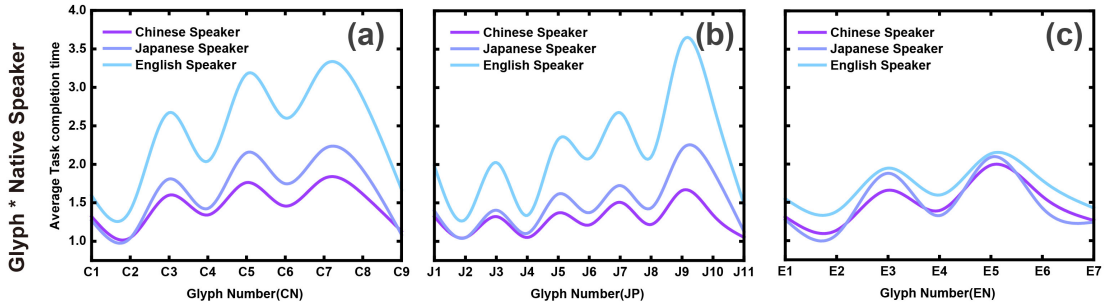



Fig. 4.7: Average Task Completion Time by Glyph and Native Language for Chinese, Japanese, and English Experiments. parts (A), (B), and (C) illustrate the interaction between Glyph and Native language groups in the Chinese, Japanese, and English experiments, respectively. In the Chinese and Japanese experiments, English native speakers generally require more time to complete tasks (parts (A) and (B)), especially for complex glyphs such as C3, C5, C7 in Chinese and J7, J9, J10 in Japanese. Under the 10 m condition, compared to 5 m, there is a notable increase in average task completion times. In the English experiment (part (C)), task completion times are similar across all native language groups, with visually similar glyphs (such as E3 and E6) requiring slightly more time than dissimilar glyphs.

### LME Model Analysis

To elucidate how complex three-way interactions influence reading performance, the most robust effects from the linear mixed-effects (LME) models are reported below:

**Chinese experiment (5 m–10 m).** English-native speakers (EL1) experienced pronounced delays when moderate-complexity characters were rendered in extreme stroke-weight serif fonts at far distances. At 5 m, the EL1 × SeB × C4 ($C > 4$) interaction incurred a +1.63 s penalty ($b = 1.63$ s, $p = .004$), and at 10 m the EL1 × SeE × C3 ($2 < C \leq 4$) term added +1.77 s ($b = 1.77$ s, $p = .036$). These results indicate that at longer distances (5 m and 10 m), both heavy-stroke (SeB) and ultra-light-stroke (SeE) serif fonts prolong task completion times for EL1 readers. Contrary to the initial hypothesis that fine-stroke serifs would facilitate recognition of complex glyphs at extreme distances, the most complex glyphs (C4, $C > 4$) instead showed increased processing time regardless of serif weight, amplifying visual parsing demands as acuity drops.

**Japanese experiment (0.5 m).** At 0.5 m, EL1 are 0.86 s slower when Glyph looks similar(J1,J3,J5,J7,J9) in the SeB font ($b = +0.86s, p = .0014$), and 1.10 s slower on Kana(J1,J2) in SeB font ($b = +1.10s, p = .009$). This suggests that, regardless of glyph type, heavy - stroke serif fonts introduce confusion and slow processing for English - native speakers in Japanese text.

**English experiment (0.5 m).** LME analysis revealed a strong effect of word length: short versus long words led to a 0.57 s faster response ($b = –0.57s, p = .0015$), medium versus long a 0.42 s advantage ($b = –0.42s, p = .020$), and mixed versus long a 0.51 s advantage ($b = –0.51s, p = .027$). Because word length varied with complexity in the English experiment (unlike the fixed lengths in the Chinese and Japanese experiments), long words such as "consideration" required a wider visual angle at the nearest viewing distance (0.5 m), resulting in significantly longer processing times, whereas shorter strings reduced visual search demands.

Taken together, the LME-derived effects, GLM results, and interaction plots suggest that VR text design is likely to benefit from a careful balance between stroke weight and glyph distinctiveness, in order to accommodate cross-lingual differences across viewing distances.

### Summary of Task Completion Time

In summary, the results for task completion time based on the hypotheses are as follows:

**Hypothesis 1: Native speakers complete tasks faster than non-native speakers.** This hypothesis is partially supported. In the Chinese experiment, the hypothesis was confirmed, as Chinese native speakers completed tasks significantly faster than non-native speakers across all distance conditions. However, in the Japanese experiment, the performance of Chinese and Japanese native speakers was comparable, with Chinese native speakers even outperforming Japanese native speakers under certain conditions (e.g., 10 m). In the English experiment, Chinese and Japanese native speakers consistently completed tasks faster than English native speakers.

**Hypothesis 2: Similar glyphs increase task completion time.** This hypothesis is fully supported. Across all language experiments and all distance conditions, both native and non-native speakers required more time to process similar glyphs, confirming the significant impact of glyph similarity.

**Hypothesis 3: Complex glyphs increase task completion time.** This hypothesis is also fully supported. In all language experiments and across all distance conditions, complex glyphs consistently required longer task completion times. Notably, in the Japanese experiment, glyph similarity had a greater impact on task completion time than glyph complexity.

**Hypothesis 4: Fonts influence task completion time.** This hypothesis is only partially supported. Overall, font type exerted minimal influence on task completion time. However, the LME analysis indicated that under specific conditions, such as the 0.5 m in the Japanese experiment and the 2.5 m distance in the Chinese experiment, the SeB font (serif bold) required significantly longer recognition times than other fonts, particularly for English-native participants.

## 4.4 Discussion of Key Findings

### 4.4.1 Differences in Visual Recognition between L1 and L2 Participants

#### Logographic L1 Participants: Faster Processing with Elevated Similarity Sensitivity

As shown in Figure 4.7(A,B,C) Native Chinese (CL1) and Japanese (JL1) participants—familiar with logographic scripts—exhibited faster task completion times in the Chinese and Japanese experiments and matched or slightly outperformed English native participants (EL1) in the English experiment. This aligns with Tan *et al.* [118], who attribute accelerated logographic processing to holistic visual strategies.

However, Figure 4.3 show that CL1 and JL1 groups incurred higher error rates for visually similar glyphs across languages. In the Chinese experiment, no significant native - language effect on error rates was found ($p > .05$), indicating comparable performance between CL1 and non - native participants , as shown in Figure 4.3(A). In contrast, native participants consistently outperformed non - natives in the Japanese and English experiments. CL1 error rates remained stable across complexity levels, as shown in Figure 4.3(B), whereas EL1 error rates increased with glyph complexity. Similarly, CL1/JL1 participants made more errors on similar alphabetic glyphs (e.g., E3, E5) than EL1 participants in the English experiment. These findings aligning with evidence that script-specific processing biases amplify sensitivity to confusable glyphs [130].

A likely explanation is that logographic readers rely on holistic, configural processing—treating each character as a unified shape—which allows rapid outline-based localization of strokes and radicals but neglects internal detail. As a result, recognition of simple and complex forms follows the same contour-driven mechanism. This strategy may underlie certain social behaviors: for example, logographic users often prefer subtitles and on-screen comments (danmaku) over dubbing, since text can be processed simultaneously with video content [138]. However, this contour-focused approach also incurs costs in the digital age: increased typographical errors and "digital dysgraphia," where users recall only rough outlines and struggle with precise stroke recall [139]. Furthermore, this bias can hinder alphabetic L2 acquisition: many Asian learners depend heavily on visuospatial-orthographic resources in short-term memory and typing tasks [140], while phonological training is often insufficient in some curricula, exacerbating difficulties with linear decoding rules [141].

#### Alphabetic L1 Participants: Complexity-Driven Slower Processing and Sensitivity

Native English (EL1) participants exhibited low error and invisibility rates for simple glyphs (Chinese C1–C2; Japanese J1–J4), but both error/invisibility rates and task completion times rose sharply once complexity exceeded $C > 0.5$ (Chinese C3–C8; Japanese

J5–J10) as shown in Figure 4.3 and Figure 4.6. In the English experiment, EL1 accuracy remained high—especially for visually similar word pairs—outperforming logographic L1 groups(CL1,JL1).

A likely explanation is that alphabetic - script speakers leverage automatic morphological decomposition—using syllable–morpheme mappings to visually parse word forms —whereas logographic - script speakers depend on holistic contour and radical matching. In tasks featuring visually similar but semantically distinct Glyphs (e.g., "consideration" vs. "conversation"), this reliance on whole - shape matching elevates error rates for logographic L1 participants. By contrast, English native speakers perform unconscious, automated decomposition without extra semantic processing cost [142]. Even experienced L2 English learners (EL2), despite years of study, tend to default to contour - matching strategies in purely shape - based tasks: they may match EL1 speeds but still incur higher error rates when glyphs share similar outlines yet differ in meaning.

On the other hand, In natural reading, English speakers decompose words into letters and letter groups, a strategy that does not generalize to multi - stroke logographs. When presented with high - complexity characters (e.g., C3, C5, C7 or J7, J9, J10) at far distances (10 m), EL1 participants could not leverage any automatic radical - or contour - based processing and thus required extra time to visually parse and compare shapes. By contrast, CL1 and JL1 have internalized radical structures and stroke patterns through years of literacy training, allowing them to chunk complex forms even in low - resolution VR conditions.

In summary, The interaction effects between glyph complexity and native language likely arise from fundamental differences in how logographic and alphabetic readers process visual forms, and combined with script - specific familiarity.Therefore EL1 participants resist word - length complexity in their native script but incur higher error rates and longer task completion times for complex logographic glyphs.

### Minimal Impact of Font on Error and Invisibility Rates

Across the three language experiments, typographic variations did not show systematic effects on error or invisibility rates in most between-group comparisons. However, linear mixed-effects (LME) analyses indicated that font characteristics did systematically modulate task completion speed under specific stimulus combinations that reached statistical significance. For instance, in the Japanese experiment, non-native speakers (EL1) exhibited marked processing delays when encountering heavy serif fonts (e.g., +0.86–1.01 s at 0.5 m viewing distance, $p < .001$), yet their error rates remained statistically indistinguishable from baseline conditions. This pattern aligns with the speed-accuracy tradeoff framework [143], where readers prioritize precision through compensatory time allocation—a strategic adaptation particularly critical in second-language (L2) processing contexts.

In conclusion, glyph similarity and complexity affect reading performance differently

depending on language backgrounds. Native speakers from the Chinese character culture (CL1 and JL1) are more influenced by visual similarity, while English native speakers (EL1) are more affected by glyph complexity in non-native scripts. Familiarity with glyph structure aids native speakers from the Chinese character culture in efficient information processing but also increases their sensitivity to similar glyphs, resulting in higher error rates. English speakers, owing to lower familiarity with non-native glyph structures, maybe require greater cognitive load to interpret complex characters. Font had a minimal effect on error and invisibility rates, influencing task outcomes only under specific experimental conditions. These findings underscore that familiarity with glyph structure enhances visual processing efficiency but also heightens sensitivity to similarity, whereas complexity emerges as a primary challenge in non-native reading.

### 4.4.2 Educational and Design Implications

This study examines how glyph complexity and similarity influence the legibility of Chinese, Japanese, and English text for native (L1) readers, second-language (L2) learners, and individuals with no prior exposure to the target language (LN). The results offer insights into how these factors affect legibility across languages and suggest directions for improving both educational strategies and interface design.

#### Learning Chinese and Japanese Characters for Native Speakers of English

The findings reveal intrinsic differences in glyph perception among native speakers of Chinese (CL1), Japanese (JL1), and English (EL1), shaped by their cultural and linguistic backgrounds. For Chinese and Japanese speakers, long-term exposure to the Chinese character system fosters high familiarity with complex and visually similar glyphs, resulting in faster task completion times. At the same time, this familiarity increases sensitivity to glyph similarity, leading to higher error rates when visually similar characters are presented.

This heightened tolerance for similarity within the structure of Chinese characters (Kanji and Hanzi) allows Chinese and Japanese speakers to overlook minor errors in visually similar characters without significantly impacting comprehension. Additionally, the inherent ambiguity in the visual features of Chinese characters has facilitated the development of fields such as handwritten character recognition [144]. Compared to English words, Chinese characters (Kanji and Hanzi) have a higher threshold for similarity, suggesting that as long as the overall structure of the character remains intact, substituting certain parts with visually similar elements—for example, simplified graphical symbols or even embedded Latin letters—may not interfere with recognition for native speakers [145]. Additionally, studies have shown that non-native beginners often rely on visual aids as a strategy for memorizing Chinese characters [146].

This tolerance for structural similarity can be leveraged to assist second language or

no prior exposure to the language learners in recognizing Chinese characters(Kanji and Hanzi). For example, educational tools could incorporate structured, hybrid glyphs that integrate familiar visual cues or supplementary English letters to help learners identify key components of complex characters. Moreover, the rich information embedded in complex characters could be utilized to create mnemonic devices or visual aids that reinforce specific character components, easing the visual perceptual load for L2 learners and supporting more efficient character recognition and learning.

**Learning English for Native Speakers of Chinese and Japanese**

For native speakers of Chinese and Japanese, English is typically learned as a second language (L2). The present findings suggest that these learners are particularly sensitive to visual similarity in English words and letters, rather than to word length or complexity per se. Unlike the logographic Chinese system, which allows for some degree of semantic encoding even with unfamiliar characters [147], the English phonetic system relies more heavily on phonological encoding and structural analysis. English word recognition depends on phonological pathways, frequency effects, morphological structure, and sequential redundancy, which together support the mapping from visual input to phonological and semantic information. Without sufficient familiarity with English phonological structures, Chinese and Japanese learners may experience additional cognitive load when processing visually similar English words or letters.

One potential strategy is to present English vocabulary in a structured, pattern-based format that leverages learners' existing familiarity with complex characters. Although some studies, such as those by Jankowski et al. [135], have explored innovative approaches like arranging English words in square formats resembling Chinese characters, this approach may come with a higher learning cost. Instead, grouping words by common prefixes, suffixes, or roots could help Chinese and Japanese learners process English words as unified, meaningful units, similar to how they interpret characters in their native languages. This method could reduce the visual perceptual load of learning English vocabulary by drawing analogies to familiar morphological structures.

Additionally, educational programs should place a greater emphasis on phonological training. Currently, much language instruction focuses on written skills, with limited exposure to spoken English outside of standardized listening assessments or media consumption. Unlike Japanese, which has a more straightforward syllabic structure, English phonemes are not always clearly linked to specific written forms.

A study focusing on enhancing Chinese character input efficiency through a multichannel approach that integrates phonetic and handwriting modalities [148] suggests that a similar multi-sensory method could also be highly beneficial in English education. Although the study specifically addresses Chinese character input, the combination of auditory, visual, and tactile channels used in this approach may prove effective in reinforcing English learning, particularly for Asian students. Therefore, future educational

tools should aim to strengthen the connections between English pronunciation, spelling, meaning, and practical usage through multi-sensory learning experiences.

#### Accessible Interface Layouts

Script-specific perceptual habits suggest the need for adaptive interface designs. Logographic users (Chinese and Japanese speakers) often tolerate or prefer denser, grid-like layouts (e.g., multi-element panels, nested icon clusters) that mirror the spatial compression of characters, whereas alphabetic-language users usually perform better with linear, minimalist layouts aligned with sequential parsing. Language-responsive layout systems could therefore dynamically adjust interface density; for example, interfaces might switch between grid-based displays (optimized for logographic users) and linear flows (for alphabetic users), enhancing information findability through spatial reorganization and typographic adaptation. Future work should establish cross-cultural design standards to better address the growing linguistic diversity of digital platforms.

#### Enhancing Language Learning and Legibility in VR Environments

VR' s immersive nature reduces external distractions and fosters sustained attention, making it an ideal platform for language study. Its multimodal interactions—combining visual, auditory, and spatial cues—can lower visual perceptual load, especially for phonology-driven languages like English, by reinforcing sound–symbol and meaning associations.

By integrating AI, a VR system can adapt in real time to each learner' s level: for L2 users, simplifying text or adding phonetic/semantic hints; for L1 users, increasing text density or complexity. VR can also recreate cultural contexts, offering contextualized practice that deepens both linguistic and pragmatic competence.

Together, immersion, multimodal support, and adaptive presentation in VR can improve text legibility and accelerate language acquisition across writing systems.

## 4.5 Chapter Summary

This chapter examined how glyph complexity and similarity affect legibility across Chinese, Japanese, and English. The study systematically investigated how these visual properties influence performance for native (L1) and non-native (L2) readers and then drew implications for design and future work.

In a VR environment using a Meta Quest 3 headset, participants from Chinese-, Japanese-, and English-language backgrounds performed a two-alternative forced-choice identification task in which a prompt glyph had to be matched to one of two candidates. Viewing distance, font, and language were systematically varied, and three outcome measures were recorded: error rate, invisibility rate, and task completion time. The

analyses examined how glyph complexity, similarity, font style, viewing distance, and native-language background jointly shaped these outcomes. Taken together, the findings provide a basis for understanding how glyph characteristics and reader background interact to shape legibility across different linguistic and cultural contexts.

The results show that for native Chinese and Japanese speakers (CL1 and JL1), familiarity with glyph structures substantially improves visual processing efficiency, enabling them to complete tasks more quickly. At the same time, this familiarity heightens sensitivity to glyph similarity, leading to higher error rates when visually confusable forms are presented. This pattern is observed not only in Chinese and Japanese but also in English, likely reflecting transfer from the native logographic background. In contrast, native English speakers (EL1) were influenced primarily by glyph complexity when reading non-native scripts (Chinese and Japanese), resulting in longer task completion times and reflecting the additional visual load required to decode unfamiliar patterns. Compared with CL1 and JL1 readers, EL1 participants were less affected by glyph similarity.

These differences further underscore the impact of linguistic and cultural background on visual perception. Characteristics of the native writing system directly influence the difficulty of processing second or unfamiliar languages, either increasing or mitigating perceptual demands. The present findings therefore highlight the importance of continued research on the visual-perceptual challenges associated with multilingual text legibility.

While this chapter establishes a foundational understanding of cross-linguistic text legibility, further work is needed to expand sample diversity and to adopt more comprehensive complexity metrics. Although font styles did not show strong effects on legibility in the present experiment, future studies should examine a broader range of typefaces, particularly those with greater stylistic variation. Such efforts could support the development of more inclusive and effective multilingual text design solutions. Ultimately, this study not only offers new perspectives for theory but also lays a practical foundation for applications in human–computer interaction, language learning, and cross-cultural communication. From the standpoint of rendering strategies, the persistent perceptual divergences observed across language groups point to the need for adaptive solutions tailored to native-script experience. By supporting multilingual accessibility, this work contributes to more inclusive XR interface infrastructures aligned with SDG 9 (Industry, Innovation, and Infrastructure).

# Chapter 5 Study on font preferences of native and non-native speakers in a virtual reality environment

Chapters 3 and 4 analyzed how viewing distance and glyph properties shape perceived readability in VR. This chapter isolates the role of *font design*, asking how style, weight, contrast, and character width influence perceived legibility under the same presentation regimes.Taken together with the distance-centric findings from Chapter 3 and the glyph-level results from Chapter 4, the present analyses constitute the third component of this dissertation' s empirical program. Chapter 6 will integrate these strands—distance, glyph properties, and font design—into a consolidated discussion of multilingual text readability in VR and translate the combined evidence into practical, script-aware guidance.

On conventional displays, typeface variables such as style, weight, contrast, and character width systematically affect legibility and recognition [131, 149–151]. For example, bolder weights can aid small-scale recognition, whereas stroke-width and internal contrast interact with word processing efficiency [131, 149, 151]. Individual differences further shape outcomes: age and vision status, as well as reading disabilities, modulate which fonts feel easiest to read [47, 152, 153]. Preferences and performance also vary with reader background; native (L1) and non-native (L2) users can show different sensitivities to font clarity and degradation [125, 154].

In VR, known font effects intersect with spatial presentation: overall sizes tend to be larger than on flat screens, motion and viewing angle can penalize fine strokes, and effective resolution imposes practical limits on usable styles [28, 49, 56, 133, 134]. Building on the distance-centric results of Chapter 3 and the glyph analyses of Chapter 4, this chapter compares *font preferences* across three orthographically distinct linguistic contexts (Chinese, Japanese, and English) and across L1/L2 reader groups. The aim is to translate established screen-typography findings [131, 149–151] into VR-appropriate guidance, taking into account population factors identified in prior work [47, 125, 152–154] and the display constraints summarized above [28, 49, 56, 133, 134]. In short, Chapter 5 focuses on how typeface style, weight, contrast, and width interact with VR presentation to shape perceived legibility and preference across languages.

## 5.1 Related work

### 5.1.1 Digital Factors Affecting Digital Reading

Previous studies on screen interfaces have extensively explored the impact of fonts on legibility. Research indicates that font characteristics such as font weight [131, 132], glyph contrast [149], and character stroke width [150, 151]significantly affect legibility. For instance, Oderkerk et al. (2021) found that bolder fonts can enhance letter recognition in small-scale reading contexts [131]. In contrast, Beier et al. (2021) suggest that bold fonts with high stroke contrast should not be prioritized in designs where letter recognition is critical [149]. Additionally, changes in character width significantly influence legibility, with wider characters improving word processing [150]. However, the difference between serif and sans-serif fonts is minimal [155, 156].

In VR environments, the principles of traditional flat design may require modification due to changes in user perspective and interaction modes. Given the spatial context, the overall font size should be slightly larger than in screen interfaces. Additionally, results can vary significantly depending on device resolution. Current research primarily focuses on factors such as device resolution, font type, font complexity, font color, font size, text background, and line spacing [49, 56, 133], as well as reading fonts under specific conditions (e.g., while moving [134] or from different angles [28]). For example, Zhou et al. found that higher resolution allows for smaller characters, but for more complex Chinese characters, larger sizes are necessary [56]).

### 5.1.2 Individual Factors Affecting legibility

While font characteristics impact individuals differently and no single font suits everyone, the primary factors influencing legibility from a personal perspective include age [47], vision [152, 157], and the presence of reading disabilities [153]. Familiarity with the text or font also plays a crucial role [158, 159].

For instance, research by Aurélie Calabrèse et al. indicates significant differences in legibility across age groups, specifically among individuals aged 8-16 years, 16-40 years, and over 40 years [47]. Increasing font size, rather than character width, has been shown to effectively aid individuals with low vision in reading [152]. For individuals with reading disabilities, sans-serif and monospaced fonts have been found to improve reading performance [153].

However, familiarity with the text or font is multifaceted. For example,readers who are familiar with specific fonts tend to read text faster. Beier, Sofie' s research found that familiar fonts enhance reading speed, illustrating the importance of exposure to particular font designs [159]. Moreover, familiarity with the content of the text can also influence legibility, as readers are better able to anticipate and recognize words and phrases they frequently encounter [158].

### 5.1.3 Cognitive Font Preferences of Native and Non-Native Speakers

Native (L1) and non-native (L2) speakers exhibit substantial differences in reading and processing visual cues due to cultural and linguistic variations. For instance, studies have shown that British participants find sans-serif fonts easier and less tiring to read, whereas Thai participants prefer serif fonts for the same reasons [154]. Moreover, using bold fonts for emphasis can aid L2 learners remember discourse content [124].

Hanna S. and colleagues compared Dutch (L1) and English (L2) reading, revealing that decreased font legibility affects L2 reading time more significantly than L1 reading time [125]. Similarly, Pae, H. et al. examined how Korean and Chinese speakers process English words using normal, alternating, and reversed fonts. They found that Chinese native speakers are more sensitive to visual distortions than Korean speakers, suggesting that the linguistic templates established in L1 play a significant role in processing English vocabulary [130].

Furthermore, the orthographic features of one's native language (L1) significantly impact English as a Second Language (ESL) learners' English acquisition. For example, learners whose native language is logographic (e.g., Chinese) may struggle with recognizing and spelling English words due to the absence of an alphabetic system in their L1. In contrast, learners from alphabetic backgrounds (e.g., French, German) tend to recognize and spell English words more efficiently due to their familiarity with alphabetic systems and spelling rules [160].

### 5.1.4 Summary

Language consists of sentences, which are composed of individual words. While research on fonts remains an evergreen topic in both traditional interfaces and VR environments, much of the existing literature has focused on font preferences among native speakers, particularly within the English language. However, Chinese and Japanese are logographic languages, whereas English is alphabetic, resulting in fundamentally different orthographic characteristics. This paper aims to bridge this gap by comparing the font preferences of native and non-native speakers across three linguistic environments: Chinese, Japanese, and English.

## 5.2 Experiment Design

This experiment examined font preferences of L1 and L2 speakers in VR. Using a Meta Quest 3 headset, participants viewed pairs of characters or words with identical content at four distances (0.5 m, 2.5 m, 5 m, and 10 m) and rated legibility on a 1–5 scale. Selections were made using the Quest Touch Pro controllers.

Thirty participants (native Chinese, Japanese, and English speakers) completed tasks in their native language and in one randomly assigned non-native language. Both quanti-

tative font ratings and qualitative user feedback were collected, and the analyses focused on how distance, glyph characteristics, native language, and font type jointly affected font preferences.

### 5.2.1 Virtual Environment

A VR corridor environment measuring 14 × 2.5 × 2.5 m was created. The corridor included a virtual text box (650 × 200 mm) supporting three language versions: Chinese, Japanese, and English. Following previous research [135], the text box background was set to black with 50% transparency, and the text was rendered in white. The text box displayed either two characters (for Chinese or Japanese) or two words (for English), with both options containing identical content.

A virtual evaluation form, rated from 1 to 5, was positioned below the user's camera view. Participants used the directional keys on the left or right Quest Touch Pro controller to select their preferred font. During evaluation, the currently selected option was highlighted in yellow, and after each selection the rating was reset to the neutral value (3). Directional lighting was used throughout the scene to ensure adequate illumination. An overview of the environment is illustrated in Figure 5.1.

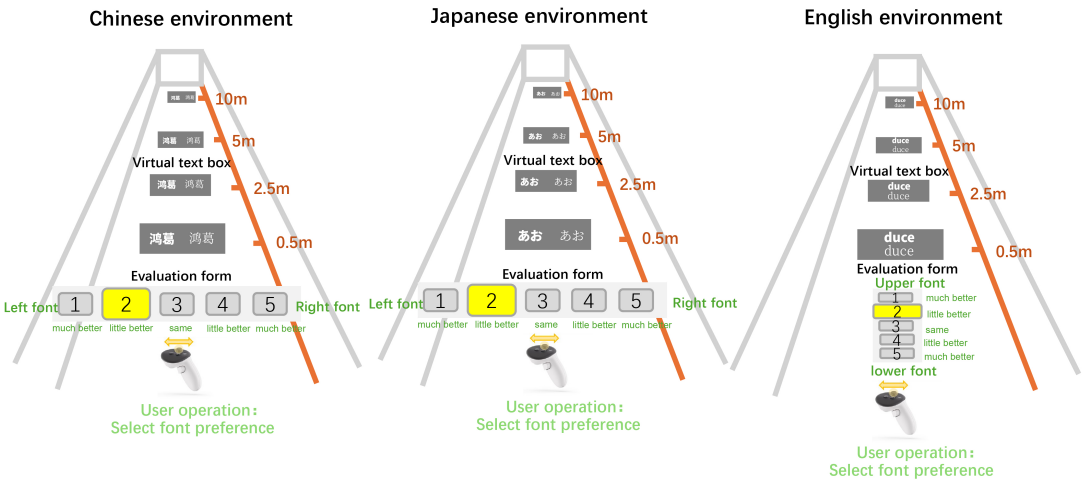



Fig. 5.1: Schematic Diagram of Experimental Setups for Chinese, Japanese, and English Tests.

### 5.2.2 Independent Variables

Participants from different native-language backgrounds were asked to select their preferred font presentation in a virtual reality environment where the same glyph (i.e., content) was displayed at a fixed distance. To systematically explore the factors that

may influence font preference, four primary independent variables were manipulated:

1. Glyph (6 groups for Japanese, 4 groups for Chinese, and 3 groups for English)
2. Fonts (6 types)
3. Distance (4 distances)
4. Native language (3 types)

### Glyph

The cross-script glyph definitions in Section 2.3.3 are followed, and the complexity-based tiers introduced in Section 2.3.3 are reused to group items in this chapter. For Chinese and Japanese, stimuli consist of isolated characters (Hanzi/Kanji and Kana); for English, stimuli are single high-frequency words. These are the smallest meaningful or visually discrete units in each script and are presented as self-contained glyphs. Participants performed no reading or comprehension task; instead, they made purely perceptual judgements of legibility and preference, so that observed effects can be attributed to font differences rather than lexical or semantic content.

According to these classifications, the glyph conditions were organized as follows:

1. Japanese: 6 groups
2. Chinese: 4 groups
3. English: 3 groups

To limit fatigue while maintaining coverage within each tier, fixed glyph sets were sampled from every category. The detailed composition of the glyph groups used in this experiment is shown in Table 5.1.

### Fonts

To minimize cross-language stylistic confounds, the same font setup as in Chapter 2 is used: harmonized variants from the Google Noto superfamily [72–77]. Noto provides consistent design across scripts (Latin, Japanese, Chinese) and has broad real-world adoption, making it suitable for cross-script comparisons.

Noto provides consistent design across scripts (Latin, Japanese, Chinese) and broad real-world adoption, making it suitable for cross-script comparisons.

Typeface factors mirror the earlier chapters: two styles (Sans-serif vs. Serif) crossed with three weights (Thin, Medium, Bold), yielding six style×weight combinations per script.

Table. 5.1: Character Glyph Classification by Feature Across Chinese, Japanese, and English

| Language | Feature | Character sets | Classification basis |
|---|---|---|---|
| Japanese | Hiragana (J1) | ぬね　びひ　おす　しむ　かち | Simple ($C \leq 0.5$) |
| | Katakana (J2) | ソン　ババ　フナ　ラガ　シイ | Simple ($C \leq 0.5$) |
| | Kanji ($C \leq 0.5$) (J3) | 山小　二又　三女　大上　才十 | Simple ($C \leq 0.5$) |
| | Kanji ($0.5 < C \leq 2$) (J4) | 呉茶　呂皇　果直　徒逆　造挑 | Normal ($0.5 < C \leq 2$) |
| | Kanji ($2 < C \leq 4$) (J5) | 格唇　渇夏　瘍棄　徳質　塊圏 | Complex ($2 < C \leq 4$) |
| | Kanji ($C > 4$) (J6) | 瞳襲　騒麗　露璽　欄覇　観覆 | Very Complex ($C > 4$) |
| Chinese | Hanzi ($C \leq 0.5$) (C1) | 乃了　一乙　入卜　工之 | Simple ($C \leq 0.5$) |
| | Hanzi ($0.5 < C \leq 2$) (C2) | 汽冉　沁由　冰决　忾汴 | Normal ($0.5 < C \leq 2$) |
| | Hanzi ($2 < C \leq 4$) (C3) | 鸦奠　鸿葛　齊疸　感庸 | Complex ($2 < C \leq 4$) |
| | Hanzi ($C > 4$) (C4) | 賽闋　孽镰　鵝騙　麋譨 | Very Complex ($C > 4$) |
| English | 0–3 letters (E1) | go he on me to up cat dog sky big duce bear green sight | Simple |
| | 4–8 letters (E2) | script crept animal beauty gallery kitchen rehabilitation disappointment | Normal |
| | over 8 letters (E3) | communication accomplishment encouragement documentation conference magnificent appreciation conference | Complex |

### Distance

Four viewing distances were defined within the virtual environment: 0.5 m, 2.5 m, 5 m, and 10 m. These distances correspond, respectively, to the minimum distance at which characters remain recognizable, a comfortable reading distance for nearsighted individuals, a suitable reading distance for farsighted individuals, and the maximum distance at which character outlines are barely distinguishable.

The purpose of setting these distances was to maximize the contrast in font preference at comfortable, near, and far distances across different fonts and glyphs.

### Native language

For the native language, following the recruitment of the native language as described in Chapter 2, this experiment also used 10 native Chinese speakers, 10 native English speakers, and 10 native Japanese speakers. The participants were also assigned to the native language and non-native language projects. Therefore, the final participants in this Chinese experiment (10 native Chinese speakers, 5 native Japanese speakers, 5 native English speakers), the participants in the Japanese experiment (10 native Japanese speakers, 5 native Chinese speakers, 5 native English speakers), and the participants in the English experiment (10 native English speakers, 5 native Chinese speakers, 5 native

Japanese speakers)

### 5.2.3 Task

Participants were presented with a series of trials in which they viewed pairs of glyph sets, each containing identical content but rendered in two different fonts. For each trial, the glyph sets were displayed at one of four fixed distances (0.5 m, 2.5 m, 5 m, or 10 m). Participants were asked to compare the two fonts and rate their relative legibility on a 1–5 scale, with 1 indicating a strong preference for the font on the left, 5 indicating a strong preference for the font on the right, and 3 representing a neutral stance. In addition to these comparisons, each participant completed the task twice— once in their native language and once in a randomly assigned non-native language. This design allowed us to systematically assess how factors such as glyph characteristics, font type, viewing distance, and native language influenced font preferences.

### 5.2.4 Procedure

Upon arrival at the experiment site, participants received a brief introduction to the purpose, procedure, and key considerations of the study. After reading and signing the informed consent form to confirm voluntary participation, they underwent brief training on how to use the VR equipment. This training included instructions on wearing the VR headset, using the Oculus controllers, and adjusting the equipment for comfort and appropriate interpupillary distance. The experiment was available in Japanese, Chinese, and English versions.

Participants selected their native language version and were then randomly assigned to one of the other non-native language versions. Each language experiment included 10 native and 10 non-native participants, with 5 participants from each of the other two language groups. Informed consent was obtained in accordance with the Institutional Review Board of Hokkaido University.

After practicing until they felt confident in understanding the requirements and mastering the procedure, the main experiment began. The character canvas and evaluation interface were presented simultaneously. The character canvas displayed two glyph sets with identical content but different fonts at one of four randomly selected distances: 0.5 m, 2.5 m, 5 m, or 10 m. Participants used the directional keys on the Oculus controller to compare and evaluate the legibility and weight of the fonts. The evaluation scale ranged from 1 to 5, with anchors from “left font is much better” to “right font is much better”. To minimize response bias and fatigue, the scale was reset to the middle position (3) after each selection.

To address potential visual fatigue from repeated evaluations, semi-structured interviews were conducted with a subset of 13 participants. The detailed interview records, with all names anonymized as numeric IDs, are available in a dedicated GitHub repos-

itory [161]. Participants were encouraged, but not required, to explain why they found one font more readable than another.

After completing the evaluation for the first language, a thank-you screen appeared and participants were given a 2–3 minute rest while the data were compiled and checked. They were then asked to complete the same task in a different non-native language. The entire experiment typically took less than 30 minutes. Finally, participants removed the VR equipment and were invited to provide brief verbal feedback on their experiences, impressions, and subjective assessments of the experiment.

The overall flow of the experiment and the detailed task steps are illustrated in Figure 5.2, which outlines both the main procedural steps and the individual tasks that participants completed during the experiment.

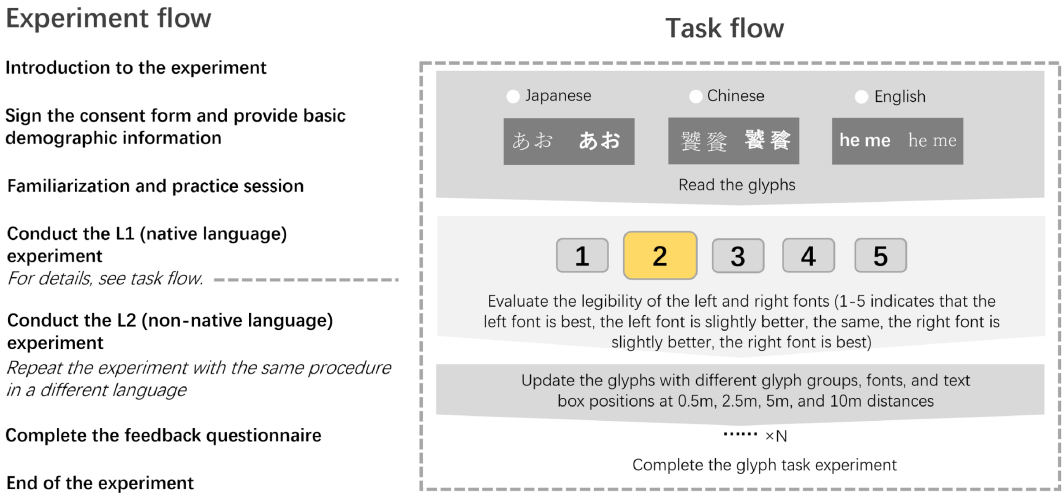



Fig. 5.2: Experiment Flow and Task Flow Diagram

### 5.2.5 Data Analysis and Processing

#### Data Collection

For each glyph presented during the trials, a comprehensive record was stored, including the following independent variables:

1. Distance (4 levels)
2. Glyphs (6 groups for Japanese, 4 groups for Chinese, and 3 groups for English)
3. Fonts (6 types)
4. Native language (3 types)

The dependent variables collected for each trial included:

1. Semi-structured interviews data (13 person in total)
2. User evaluation results(Will be converted to the score of each font in terms of distance, font, and native speaker conditions)

In total, 12,476 valid trial records were obtained: 5,802 for the Japanese experiment (6 glyph groups × 4 distances × 6 fonts × 20 participants), 3,834 for the Chinese experiment (4 glyph groups × 4 distances × 6 fonts × 20 participants), and 2,840 for the English experiment (3 glyph groups × 4 distances × 6 fonts × 20 participants). Although the experimental design anticipated these exact counts, small discrepancies between expected and actual numbers arose due to the random-combination procedure and unavoidable system-related factors during data collection. The distribution of all variables nevertheless remained nearly uniform, ensuring that these minor deviations did not compromise the overall balance or representativeness of the dataset. In addition, qualitative data from the interviews were incorporated to supplement and interpret the quantitative findings, providing a more comprehensive picture of participants' performance and cognitive processes during the experiment.

**Data Pre-processing**

Because stimuli were randomized within a full-factorial design, minor frequency imbalances arose across font-comparison conditions. To obtain comparable metrics, a normalized scoring system was applied that translated the 5-point evaluation scale into comparative scores for each font. Participants' ratings were converted as follows:

- 1 : 2 points for the left font
- 2 : 1 point for the left font
- 3 : no points for either font
- 4 : 1 point for the right font
- 5 : 2 points for the right font

This trichotomous scoring system (0, 1, or 2 points per comparison) operationalized three distinct preference levels:

- 0 point : Absolute neutrality or participants chosed another font.
- 1 point : This font is little better than the other font
- 2 points : This font is very good than the other font

To disentangle the effects of font preference, glyph, distance, and native language, the analysis proceeded in four investigative phases.

First, General Linear Model (GLM) analyses were run on the full dataset and separately for each distance (0.5 m–10 m) to test main effects (Font, Distance, Glyph Complexity, Native Language) and interactions (e.g., Font × Distance, Font × Glyph, Font × Distance × Native).

Next, for factors showing significant effects, such as Distance × Font, results were visualized using error-bar plots (displaying mean values with 95% confidence intervals), and univariate tests with estimated marginal means (EMMs) and Tukey' s HSD post-hoc comparisons were conducted for each group to examine trends in font styles (sans vs. serif) and font weights across distances.

Then, for more detailed combinations, such as Font × Distance × Glyph and Font × Distance × Native, error-bar plots with 95% confidence intervals were again used to inspect trends both within groups and overall. The score distribution was highly skewed: 65.1% of scores were 0 points, 26.1% were 1 point, and only 8.8% were 2 points. Because most font scores were concentrated at the lower end, the overall mean was relatively low. To better capture the underlying patterns, the y-axis of the plots was rescaled to highlight subtle differences among groups. Subsequent analyses included ANOVA and Tukey' s HSD post-hoc tests for each group to evaluate statistical significance.

Finally, to present the findings more clearly, winning rates (the number of times a font was selected divided by the total number of trials) were recalculated for font–condition combinations showing notable trends and converted into probabilities, facilitating clearer comparison and visual interpretation.

## 5.3 Results

### 5.3.1 Semi-structured Interview Analysis

During the experiment, some early participants (IDs 02, 05, 06, 10; see [161]) reported distraction and boredom arising from the repetitive nature of the selection tasks. To address potential cognitive fatigue induced by these repeated choices, semi-structured interviews were conducted with a targeted subset of 13 participants during the mid-to-late stages of the study. The primary objective of these interviews was to obtain deeper insight into participants' cognitive processes, subjective experiences, and any challenges they encountered throughout the experiment.

The interview questions were designed to explore several key areas, including, but not limited to:

- Personal Preference: Do you have any specific fonts that you prefer in your daily life?
- Reason for Choice: What makes the current font more readable to you?

- Emotional Response: How does this font affect your feelings or emotional response?
- Challenges: What difficulties or challenges did you experience during the experiment?

These interviews were systematically recorded and transcribed, followed by a thematic analysis to identify recurring themes and patterns. The analysis revealed that, while legibility was the primary concern for most participants—consistent with the central objective of the experimental design—some participants were also influenced by emotional preferences, such as a liking for thin or bold fonts. Additionally, a small subset of participants prioritized aesthetic appeal over legibility in their font choices.

The thematic analysis of the interview data highlighted several critical insights:

**Language-Specific Differences:**

- Chinese and Japanese: At the closest distances (0.5 m), participants generally found serif fonts more readable than sans-serif fonts, although overly bold fonts were considered inappropriate. As the viewing distance increased (2.5 m to 10 m), the influence of font type diminished, while the impact of font weight became more pronounced. For complex glyphs($2 < C \leq 4$,$C > 4$), thinner fonts were preferred for legibility, whereas for simpler glyphs, bolder fonts were favored.
- English: Thin sans-serif fonts were found to be more readable at close distances (0.5 m), while thin serif fonts were preferred at a comfortable near distance (2.5 m). At the farthest distances (10 m), bold fonts were easier to read. One participant noted that short words were readable in any font, but longer words were more legible in thinner fonts.

**Font Aesthetics vs. legibility:**

- Close Distances: At the nearest and comfortable near distances (0.5 m and 2.5 m), participants showed a preference for font aesthetics. For Chinese and Japanese text, serif fonts were more popular, whereas for English text, sans-serif fonts were favored. Font weight preferences varied depending on the complexity of the glyphs.
- Far Distances: As the distance increased (5 m and 10 m), the emphasis shifted from aesthetics to legibility. Participants exhibited a growing preference for sans-serif fonts for Chinese and Japanese text, and for serif fonts for English text.

The interview results provide valuable background information and insights that supplement and help in interpreting the experimental results.

### 5.3.2 Analysis of Individual Factors

The experiment was conducted across three language environments. Participants with different native languages completed tasks in both their native language and one non-native language. In each language condition, six distinct fonts were evaluated under varying viewing distances and levels of glyph complexity. Overall, as shown in Figure 5.3, English fonts received the highest scores, whereas Chinese fonts received the lowest. Pairwise comparisons revealed significant differences (CN vs. EN: $MD = -0.22, p < .001$; CN vs. JP: $MD = -0.05, p < .001$; EN vs. JP: $MD = 0.17, p < .001$).

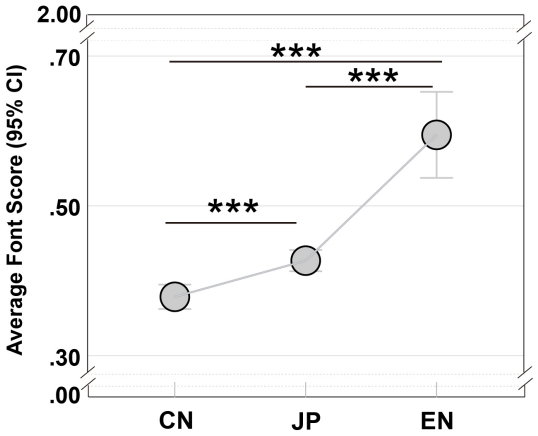



Fig. 5.3: Overall Font Preferences of Chinese, Japanese, and English Experiments.The x-axis represents different language experiments (CN = Chinese, JP = Japanese, EN = English), and the y-axis shows the mean font scores (with 95% confidence intervals). Significant differences are marked by "*" ($*p < 0.05, **p < 0.01, ***p < 0.001$).

In experimental design, two fonts were presented in random pairs, and participants were asked to choose the more legible font under the given conditions. The selected font received either 1 or 2 points, whereas the non-selected font received 0 points. If participants judged the two fonts to be equivalent, both fonts were assigned 0 points. Consequently, the overall results can be interpreted in two ways. First, the influence of font characteristics on legibility may vary by language: English fonts received higher scores compared to Chinese or Japanese fonts. Second, participants' performance on non-native language tasks may have had a strong influence. Specifically, since nearly all non-native speakers (CL1 and JL1) were L2 speakers in the English experiment, while the non-native groups in the Japanese and Chinese experiments comprised L2 participants, preferences for unfamiliar scripts may differ substantially from those of native speakers —particularly in the Chinese experiment.

Next, between-subjects tests were conducted on both the main effects (Font, Distance, Glyph, and Native Language) and their interaction terms for each language experiment (CN, EN, JP), as summarized in Table 5.2. Although the main effects showed high

levels of significance, the primary interest lay in more complex interaction patterns (i.e., two-way and three-way interactions). Accordingly, the discussion below emphasizes group-level effects of these interaction terms across the different language conditions.

Table. 5.2: Tests of Between-Subjects Effects by Glyph, Font, Distance, and Native Language

| Experiment | Case | Sum of Squares | df | Mean Square | F | Sig. | partial $\eta^2$ |
|---|---|---|---|---|---|---|---|
| Chinese | Font | 40.134 | 5 | 8.027 | 27.557 | 0.000 | 0.030 |
| | Distance | 49.753 | 3 | 16.584 | 56.936 | 0.000 | 0.037 |
| | Glyphs | 0.744 | 4 | 0.186 | 0.638 | 0.635 | 0.001 |
| | Native | 0.815 | 2 | 0.408 | 1.399 | 0.247 | 0.001 |
| | Native × Font | 19.101 | 10 | 1.910 | 6.558 | 0.000 | 0.015 |
| | Font × Distance | 27.779 | 15 | 1.852 | 6.358 | 0.000 | 0.021 |
| | Font × Glyphs | 24.751 | 20 | 1.238 | 4.249 | 0.000 | 0.019 |
| | Native × Font × Distance | 13.787 | 30 | 0.460 | 1.578 | 0.024 | 0.011 |
| | Native × Font × Glyphs | 17.353 | 40 | 0.434 | 1.489 | 0.025 | 0.013 |
| | Font × Distance × Glyphs | 21.616 | 60 | 0.360 | 1.237 | 0.105 | 0.016 |
| Japanese | Font | 79.568 | 5 | 15.914 | 44.090 | 0.000 | 0.030 |
| | Distance | 13.368 | 3 | 4.456 | 12.345 | 0.000 | 0.005 |
| | Glyph | 3.918 | 7 | 0.560 | 1.551 | 0.145 | 0.002 |
| | Native | 36.253 | 2 | 18.126 | 50.220 | 0.000 | 0.014 |
| | Distance × Font | 28.031 | 15 | 1.869 | 5.177 | 0.000 | 0.011 |
| | Glyphs × Font | 37.325 | 35 | 1.066 | 2.955 | 0.000 | 0.014 |
| | Native × Font | 36.654 | 10 | 3.665 | 10.155 | 0.000 | 0.014 |
| | Distance × Glyphs × Font | 36.598 | 105 | 0.349 | 0.966 | 0.581 | 0.014 |
| | Distance × Native × Font | 25.792 | 30 | 0.860 | 2.382 | 0.000 | 0.010 |
| | Glyphs × Native × Font | 37.676 | 70 | 0.538 | 1.491 | 0.005 | 0.014 |
| English | Font | 120.156 | 5 | 24.031 | 51.059 | 0.000 | 0.089 |
| | Distance | 2.141 | 3 | 0.714 | 1.516 | 0.208 | 0.002 |
| | Glyph | 3.899 | 2 | 1.950 | 4.142 | 0.016 | 0.003 |
| | Native | 15.700 | 2 | 7.850 | 16.679 | 0.000 | 0.013 |
| | Font × Distance | 52.596 | 15 | 3.506 | 7.450 | 0.000 | 0.041 |
| | Font × Glyphs | 2.433 | 10 | 0.243 | 0.517 | 0.879 | 0.002 |
| | Font × Native | 13.097 | 10 | 1.310 | 2.783 | 0.002 | 0.010 |
| | Font × Distance × Glyphs | 17.249 | 30 | 0.575 | 1.222 | 0.189 | 0.014 |
| | Font × Distance × Native | 27.148 | 30 | 0.905 | 1.923 | 0.002 | 0.022 |
| | Font × Glyphs × Native | 10.258 | 20 | 0.513 | 1.090 | 0.353 | 0.008 |

In the Chinese group, significant interactions were observed for:

- Font × Distance ($F = 6.358$, $p < 0.001$, $\eta^2 = 0.021$)
- Font × Glyphs ($F = 4.249$, $p < 0.001$, $\eta^2 = 0.019$)
- Font × Native ($F = 6.558$, $p < 0.001$, $\eta^2 = 0.015$)
- Native × Font × Distance ($F = 1.578$, $p < 0.05$, $\eta^2 = 0.011$)
- Native × Font × Glyphs ($F = 1.489$, $p < 0.05$, $\eta^2 = 0.013$)

In the Japanese group, significant interactions included:

- Font × Distance ($F = 5.177$, $p < 0.001$, $\eta^2 = 0.011$)

- Font × Glyphs ($F = 2.955$, $p < 0.001$,$\eta^2 = 0.014$)
- Native × Font ($F = 10.155$, $p < 0.001$, $\eta^2 = 0.014$)
- Native × Font × Distance ($F = 2.382$, $p < 0.05$, $\eta^2 = 0.010$)
- Glyphs × Native × Font ($F = 1.491$, $p < 0.05$, $\eta^2 = 0.014$)

For the English group, significant interactions were found for:

- Font × Distance ($F = 7.450$, $p < 0.001$, $\eta^2 = 0.041$)
- Nation × Font ($F = 2.783$, $p < 0.01$, $\eta^2 = 0.010$)
- Font × Distance × Native ($F = 1.923$, $p < 0.01$, $\eta^2 = 0.022$)

Notably, in the English experiment, the interaction between Font and Glyphs did not reach statistical significance ($p > 0.05$), nor did any other interactions.

Although some of these interaction effects exhibit medium effect sizes, the complexity of the interactions—combined with the sensitivity of font evaluation—requires careful, context-dependent interpretation. Accordingly, the experiments were organized by language, and detailed comparisons of parameter combinations were carried out within each language environment.

**Chinese Experiment**

Figure 5.4 illustrates the overall comparative relationships among fonts in the Chinese experiment at different distances. The analysis first considers these global patterns. For more detailed effects—especially the complex interactions of Font × Native and Font × Glyph at each distance—separate GLM models were fitted for each distance to examine between-group differences. Whenever a complex interaction effect reached significance at a given distance, a univariate test on the interaction term was first conducted to confirm an overall group effect (Univariate Tests)—for example, when the SaB font yielded a significant result with a non-negligible effect size ($\eta^2 = 0.023, p < .05$)—and estimated marginal means (EMMs) were then used for pairwise comparisons. For instance, under the SaB condition, native English speakers (EL1) showed greater acceptance of that font than native Japanese speakers (JL1) (EL1 > JL1, $p < .05$).

To maintain readability, only the direction of the mean difference (e.g., "EL1 > JL1") and the associated $p$-value are reported in the text, and each comparison is interpreted in terms of its practical relevance for VR text legibility. This procedure makes it possible to retain broad trends while still offering finer-grained case analyses where informative.

Based on the between-group effects at each distance, significant Font × Glyph interactions were observed at 0.5 m and 2.5 m $\big(F(15, 1010) = 1.97, p < .05, \eta^2 = 0.031$ at 0.5 m; $F(15, 974) = 1.97, p < .001, \eta^2 = 0.046$ at 2.5 m$\big)$, along with significant Font × Native interactions $\big(F(10, 1010) = 3.527, p < .001, \eta^2 = 0.036$ at 0.5 m; $F(10, 1010) =$

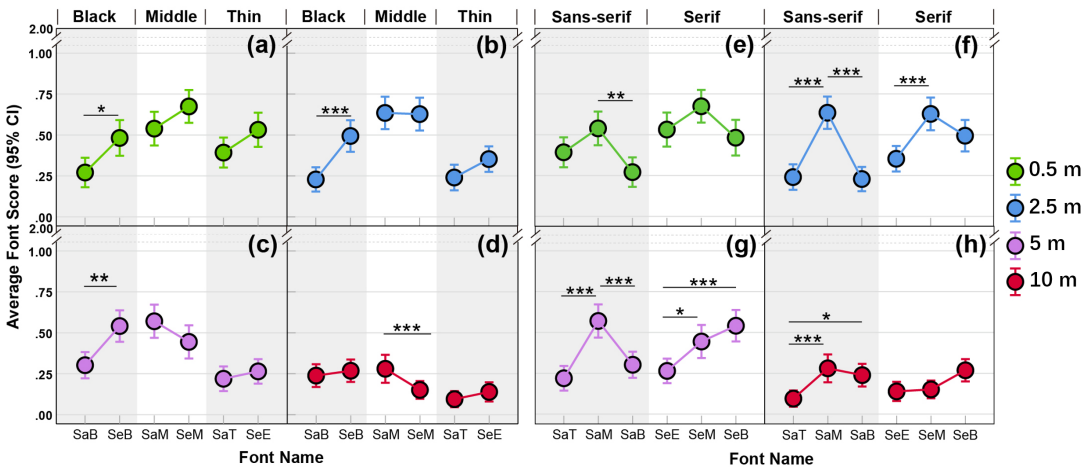



Fig. 5.4: Legibility comparisons across Chinese font styles and weights at four viewing distances. Panels (a–d) present serif vs. sans-serif contrasts under each weight category, while panels (e–h) present weight contrasts (Black, Middle, Thin) within each font style. The x-axis lists font names, and the y-axis shows mean performance scores (0–2) with 95% confidence intervals. Post-hoc pairwise comparisons (EMM) are indicated by asterisks ($*p < .05$, $**$ $p < .01$, $***p < .001$).

$2.41, p < .01, \eta^2 = 0.026$ at 2.5 m). EMM contrasts for 0.5 m and 2.5 m are therefore examined in more detail. In contrast, at 5 m and 10 m the between-group effects of these interaction terms were not significant ($p > .05$), and detailed contrasts are not reported for those distances.

**Serif Font vs. Sans-Serif Font**

As shown in Figure 5.4 (panels a–d), for 0.5 m–5 m (panels a–c) the bold serif font (SeB) tended to yield higher scores than the bold sans-serif font (SaB). At 0.5 m, JL1 and CL1 showed greater acceptance of SeB compared with EL1 ($p < .01, \eta^2 = 0.01$; JL1 > EL1, $p = .01$; CL1 > EL1, $p < .01$). At 10 m (panel d), although all fonts performed poorly overall, the medium-weight sans-serif font (SaM) outperformed the medium-weight serif font (SeM).

**Font Weight Comparisons**

As shown in Figure 5.4 (panels e–h), at close distances 0.5 m and 2.5 m (panels e–f) the medium weights (SaM, SeM) achieved the highest scores. However, for these medium weights neither the Font × Glyph nor the Font × Native interaction reached significance ($p > .05$), and no further pairwise contrasts are reported for them. Although SaT and SeE showed slightly better performance at 0.5 m, from 2.5 m onward they lagged behind the medium and bold weights. According to the EMM results, at 0.5 m and 2.5 m within SaT and SeE the more complex glyphs tended to receive higher scores. For example, at 0.5 m in SaT the simple tier C1 was rated lower than C2, C3, and C4 ($p < .01$), and

in SeE C1 was rated lower than C3 ($p < .05$) and C4 ($p < .01$). At 2.5 m in SaT, C1 and C2 were rated lower than C3 and C4 (all $p < .01$). By contrast, at 2.5 m in SeB, simpler glyphs were favored ($p < .05, \eta^2 = 0.01$; C1 > C2, $p < .05$; C1 > C3, $p < .01$; C1 > C4, $p = .01$).

At far distances 5 m–10 m (panels g–h), particularly at 5 m, SaM and SeB yielded the highest scores. Although 10 m scores were generally low, the trend across weights broadly mirrored that of 5 m—SaM outperformed SaT, and serif fonts showed increasing scores from light to bold. Because the interaction effects at 5 m and 10 m were not significant ($p > .05$), no additional specific comparisons are reported for those distances.

In summary, for Chinese characters at close viewing distances (0.5 m–2.5 m), the bold serif font (SeB) appears most legible overall—especially for JL1 and CL1—and simple glyphs pair best with SeB, whereas complex glyphs benefit more from lighter weights such as SaT or SeE. One interpretation is that at close range the visual system prefers a balance between glyph complexity and font weight: highly complex characters require lighter weights to avoid visual overload, whereas very simple characters benefit from slightly heavier weights that enhance clarity. In addition, close viewing makes fine details more discernible, so decorative strokes in simple glyphs may confer greater comfort for native readers (CL1 and JL1).

**Japanese Experiment**

Figure 5.5 illustrates the overall font performance in the Japanese experiment at four viewing distances. The subplot layout and parameter settings match those of the Chinese experiment. As in the Chinese analysis, global font effects are first assessed from Figure 5.5, and then separate GLM models are fitted for each distance to test Font × Native, Font × Glyph, and Font × Native × Glyph interactions. Significant interactions are followed up with Univariate Tests and EMM-based post-hoc contrasts, using the same data structure described in Section 5.3.2.

Based on between-group tests of interaction terms at each distance:

- 0.5 m: Font × Native was significant ($F(10, 1438) = 4.388,\ p < .001,\ \eta^2 = 0.032$), as was Font × Glyph × Native ($F(50, 1438) = 1.564,\ p < .01,\ \eta^2 = 0.056$).
- 2.5 m: Font × Native ($F(10, 1450) = 2.359,\ p < .01,\ \eta^2 = 0.017$) and Font × Glyph ($F(25, 1450) = 2.180,\ p < .001,\ \eta^2 = 0.039$) were significant.
- 5 m: Font × Native ($F(10, 1468) = 4.319,\ p < .001,\ \eta^2 = 0.031$), Font × Glyph ($F(25, 1468) = 2.138,\ p < .001,\ \eta^2 = 0.038$), and Font × Glyph × Native ($F(50, 1468) = 1.405,\ p < .05,\ \eta^2 = 0.049$) were all significant.
- 10 m: Font × Glyph was significant ($F(25, 1446) = 1.638,\ p < .05,\ \eta^2 = 0.030$).

**Serif Font vs. Sans-Serif Font**

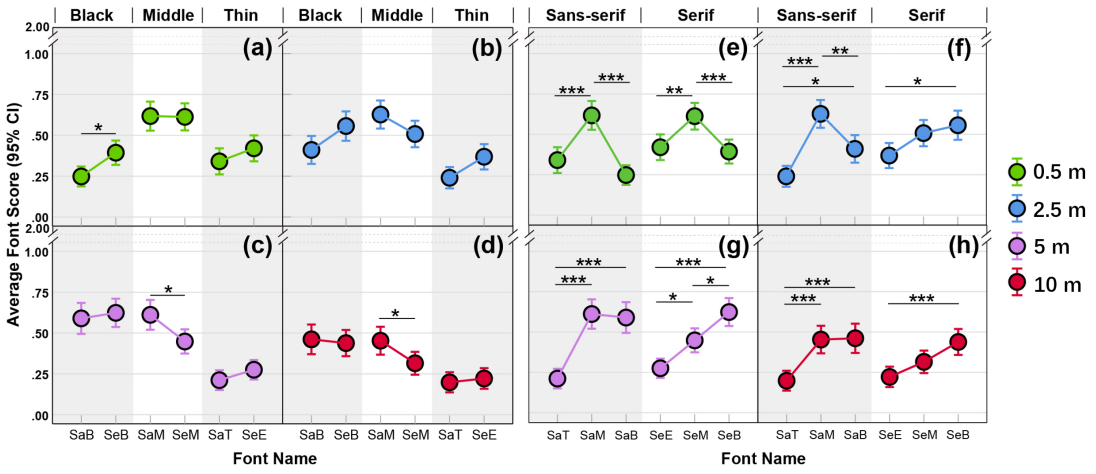



Fig. 5.5: Legibility comparisons across Japanese font styles and weights at four viewing distances. Panels (a–d) present serif vs. sans-serif contrasts under each weight category, while panels (e–h) present weight contrasts (Black, Middle, Thin) within each font style. The x-axis lists font names, and the y-axis shows mean performance scores (0–2) with 95% confidence intervals. Post-hoc pairwise comparisons (EMM) are indicated by asterisks ($*p < .05$, $**$ $p < .01$, $***p < .001$).

At 0.5 m (panel a), SeB scored significantly higher than SaB. EMM comparisons showed that Japanese native speakers (JL1; $p < .001$, $\eta^2 = 0.051$) preferred SeB over SaB (SeB > SaB, $p < .001$). In the three-way Font × Native × Glyph contrasts, JL1 at glyph J1 ($p < .001$, $\eta^2 = 0.028$) showed an even stronger SeB preference (SeB > SaB, $p < .01$).

By 5 m and 10 m (panels c–d), SaM outperformed SeM, especially for glyph J3 (5 m: $p < .001$, $\eta^2 = 0.039$; 10 m: $p < .05$, $\eta^2 = 0.023$) (5 m: SaM > SeM, $p < .001$; 10 m: SaM > SeM, $p < .05$). English native speakers (EL1) at 5 m ($p < .001$, $\eta^2 = 0.042$) also significantly favored SaM over SeM (SaM > SeM, $p < .001$). This pattern suggests that decorative serifs enhance legibility at close range but introduce visual clutter at greater distances.

**Font Weight Comparisons**

Panels e–h of Figure 5.5 compare light, medium, and bold weights of each font.

At 0.5 m (panel e), medium weights (SaM, SeM) significantly outperformed both light and bold weights. For example, at glyph J4 ($p < .001$, $\eta^2 = 0.020$), SaM outperformed SaT and SaB (SaM > SaT, $p < .001$; SaM > SaB, $p < .01$).

At 2.5 m (panel f), within sans-serif fonts, SaM remained the best, but SaB began to pull ahead of SaT, especially for J3 ($p < .001$, $\eta^2 = 0.014$; SaB > SaT, $p < .001$). Within serif fonts, SeB performed better than SeE and SeM, particularly for J1 ($p < .001$, $\eta^2 = 0.015$; SeB > SeE, $p = .001$; SeB > SeM, $p < .05$). Because Hiragana (J1) consists

of simple curves, a decorative bold weight may enhance recognition without hindering cognition. In addition, JL1 preferred SeB more than CL1 and EL1 (JL1 > CL1, $p < .001$; JL1 > EL1, $p < .01$).

By 5 m–10 m (panels g–h), both SaM and SaB outperformed SaT across most glyphs. At 5 m (panel g), SaT was weaker than both SaB and SaM (J1–J5: SaB > SaT, $p < .05$; J2–J5: SaM > SaT, $p < .05$), although no significant contrasts emerged for the most complex glyph J6 ($p > .05$). Additionally, JL1 accepted SaB and SaM more than CL1 (SaB: JL1 > CL1, $p < .001$; SaM: JL1 > CL1, $p < .05$), but less than EL1 for SaB (JL1 < EL1, $p < .01$). At 10 m, only for the simplest glyphs did SaB and SaM significantly outperform SaT (J1–J3: SaB > SaT, $p < .05$; J2–J3: SaM > SaT, $p < .05$). Serif weights followed a similar trend: SeB was strongest. At 5 m, especially for J3 ($p < .001$, $\eta^2 = 0.039$), SeB outperformed SeM and SeE (SeB > SeM, $p < .01$; SeB > SeE, $p < .001$), and, as with sans-serif fonts, JL1 preferred SeB more than CL1 (JL1 > CL1, $p < .01$). At 10 m, SeB significantly exceeded SeE and SeM for J2–J3 (J2: $p < .01$, $\eta^2 = 0.019$; J3: $p < .05$, $\eta^2 = 0.023$), with pairwise contrasts indicating SeB > SeE and SeB > SeM ($p < .01$ and $p < .05$, respectively).

Overall, these results indicate that at close distances bold decorative weights can improve the legibility of simple glyphs, whereas at far distances light weights become unsuitable. Japanese native readers (JL1) tend to favor bold weights at near ranges, English native readers (EL1) consistently show strong preferences for sans-serif fonts across distances, and Chinese native readers (CL1) are comparatively less inclined toward bold weights.

### English Experiment

Figure 5.6 illustrates the overall font performance in the English experiment at four viewing distances, with subplot arrangement and analysis procedures matching those of the Chinese experiment. Global font effects are first assessed from Figure 5.6, and then separate GLM models are fitted for each distance to test Font × Native and Font × Glyph interactions. Significant interactions are followed up with Univariate Tests and EMM-based post-hoc contrasts, using the same data structure described in Section 5.3.2.

**Serif Font vs. Sans-Serif Font**

In Figure 5.6 (panels a–d), although no post-hoc comparison between serif and sans-serif fonts of the same weight reached statistical significance, clear trends emerged. At 0.5 m and 2.5 m, the light serif font (SeT) scored lower than its sans-serif counterpart (SaT). At 5 m and 10 m, the bold serif font (SeB) scored lower than the bold sans-serif font (SaB), and the medium serif (SeM) scored lower than the medium sans-serif (SaM).

At 10 m, the Font × Native interaction was significant ($F(10, 718) = 2.198$, $p < .05$, $\eta^2 = 0.032$), and EMM comparisons indicated that Japanese native speakers (JL1) accepted SeM more than Chinese (CL1) and English (EL1) natives ($p < .01$, $\eta^2 = 0.016$; JL1 > CL1, $p = .001$; JL1 > EL1, $p < .05$). Overall, however, sans-serif fonts slightly

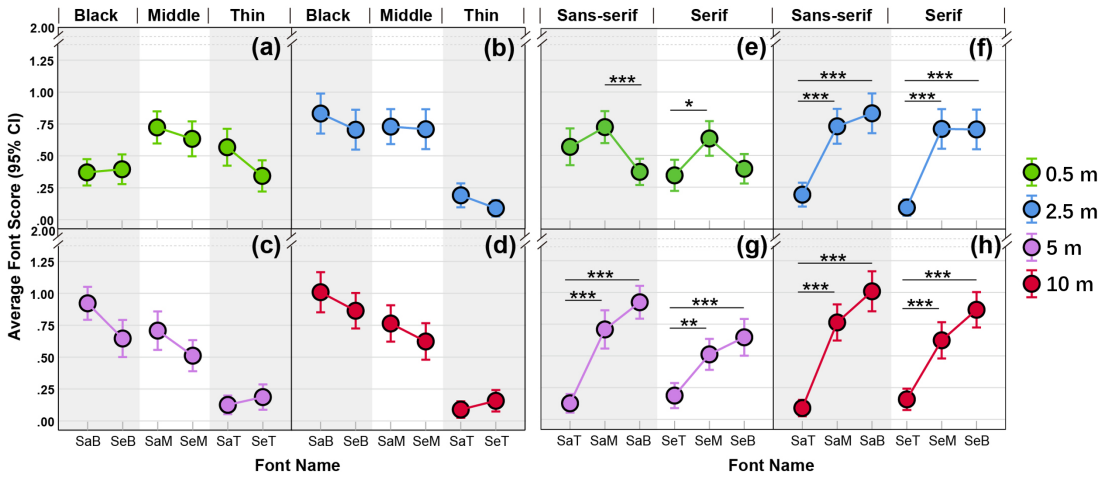



Fig. 5.6: Legibility comparisons across English font styles and weights at four viewing distances. Panels (a–d) present serif vs. sans-serif contrasts under each weight category, while panels (e–h) present weight contrasts (Black, Middle, Thin) within each font style. The x-axis lists font names, and the y-axis shows mean performance scores (0–2) with 95% confidence intervals. Post-hoc pairwise comparisons (EMM) are indicated by asterisks ($*p < .05$, $**$ $p < .01$, $***p < .001$).

outperformed serif fonts.

**Font Weight Comparisons**

In Figure 5.6 (panels e–h), multiple post-hoc tests revealed significant effects. At the closest distance 0.5 m, medium weights (SaM and SeM) performed best for both serif and sans-serif fonts. Because no detailed interaction effects reached significance at 0.5 m ($p > .05$), EMM contrasts were not examined further for that distance.

From 2.5 m through 10 m (panels f–h), the same pattern held: both medium and bold weights outperformed light weights, and bold weights gained relative advantage as distance increased. In the between-group analyses, the Font × Native interaction was significant at 2.5 m ($F(10, 670) = 2.545$, $p < .01$, $\eta^2 = 0.040$) and at 10 m ($F(10, 718) = 2.198$, $p < .05$, $\eta^2 = 0.032$). EMM results showed significant native-group differences for SaB at 2.5 m ($p < .001$, $\eta^2 = 0.028$), and for SeB and SeM at 2.5 m ($p < .01$, $\eta^2 = 0.016$ and $\eta^2 = 0.018$, respectively). Specifically, at 2.5 m, JL1 rated SaB, SeB, and SeM higher than CL1 and EL1 (SaB: JL1 > CL1, $p < .01$; JL1 > EL1, $p < .001$; SeB: JL1 > CL1, $p < .05$; JL1 > EL1, $p < .01$; SeM: JL1 > CL1, $p < .05$; JL1 > EL1, $p < .01$).

At 10 m, JL1 accepted SaB, SeB, and SeM more than CL1 (SaB: $p < .01$; SeB: $p < .001$; SeM: $p = .001$) and accepted SeB and SeM more than EL1 (SeB: $p < .001$; SeM: $p = .001$).

In summary, sans-serif fonts slightly outperformed serif fonts overall. At the nearest distance, medium-weight fonts held a clear advantage. As distance increased, both

medium and bold weights remained strong—especially bold weights—which were most preferred by Japanese native readers. Light weights (both serif and sans-serif) were generally unsuitable for VR presentation at comfortable (2.5 m–5 m) or far (10 m) distances.

**Summary of Individual Factors**

Because the number of possible interaction combinations is large, the main text reports only those cases in which all three tests—between-group effects, Univariate Tests, and EMM pairwise comparisons—were significant and the effect size was non-negligible. While this approach guarantees statistical rigor, it can obscure overall trends and make it harder to visually assess the full data pattern. Since the Font × Native and Font × Glyph interactions were most consistently significant across distances, these two interaction types are visualized in Figures 5.7, 5.8, and 5.9, with each significant contrast annotated by "*", "**", or "***". This serves two purposes: (1) to display the complete set of results for an intuitive view of global trends, and (2) to highlight the significant comparisons for detailed inspection. To enhance readability, Figures 5.7, 5.8, and 5.9 present the detailed contrasts, with significant effects indicated directly in the plots. This summary subsection integrates these visual trends with the quantitative findings across the Chinese, Japanese, and English experiments.

For the Chinese experiment, findings suggest that at extreme long distances (10 m) serif fonts and light fonts (e.g., SaT, SeM, and SeE) should be avoided (see panels 3, 5, and 6 in Figure 5.7). Moreover, complex glyphs—whether rendered in bold or light weights—resulted in suboptimal legibility (refer to panels 1–6 in Figure 5.7). Therefore, it is recommended to increase font size when possible, particularly for native English speakers (see panels 8–11 in Figure 5.7). For simple Chinese characters, increasing font weight at longer distances improves legibility for both native and non-native groups. Conversely, at closer distances (0.5 m-2.5 m), the bold serif font SeB demonstrates superior acceptability, especially among Japanese (JL1) and Chinese (CL1) native speakers. Specifically, simple glyphs benefit from bold serif fonts (e.g., SeB), while complex glyphs perform better with light fonts like SaT or SeE. This likely reflects the eye's need for balanced weight distribution: lighter fonts reduce visual density for complex characters, while slightly heavier weights optimize simple characters (see panels 2, 3, 5, and 6 in Figure 5.7). Additionally, decorative serif details enhance comfort for Chinese natives and Japanese natives at close range. These aesthetic considerations should be balanced against legibility requirements.

For the Japanese experiment, the trend of Chinese characters (Kanji: J3–J6) was similar to that observed in the Chinese experiment. At extreme long distances (10 m), light fonts such as SaT and SeE are not recommended (see panels 3 and 6 in Figure 5.8). Furthermore, for complex glyph forms, the use of bold fonts (e.g., SaB and SeB) should be minimized (see panels 1 and 4 in Figure 5.8). Instead, medium-weight fonts—adjusted in

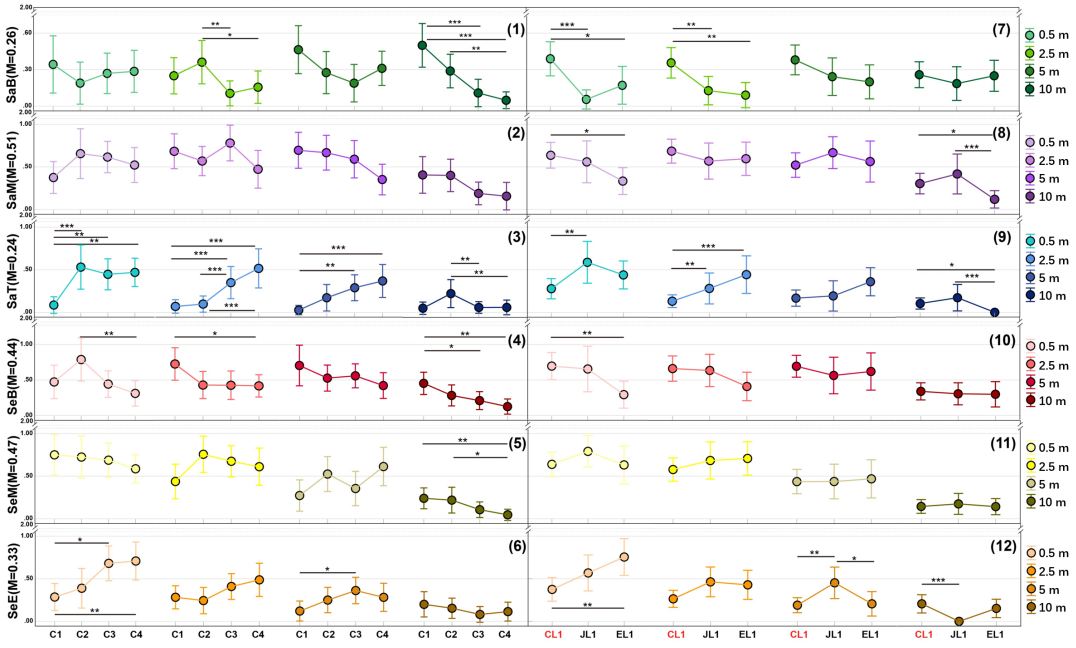



Fig. 5.7: Font Performance Trends of Chinese Fonts Under Varying Glyph Complexity, Native Language Groups, and Viewing Distances. Panels (1)–(6) illustrate how font performance varies with different glyph complexities (x-axis), while Panels (7)–(12) show how performance differs across native language groups (x-axis). In all panels, the y-axis indicates the mean font performance scores (0–2 scale) with 95% confidence intervals, where "M" denotes the average score for each font. Separate post hoc analyses were conducted for each trend, with significant differences marked by "*" ($*p < 0.05, **p < 0.01, ***p < 0.001$).

size according to complexity—are recommended. For simple glyphs, a moderate weight increase enhances legibility, with decorative bold fonts improving perception at close range. Crucially, native language groups exhibit distinct preferences: English speakers (EL1) favor sans-serif fonts overall, Chinese speakers (CL1) show lower acceptance of bold fonts, while Japanese speakers (JL1) demonstrate strong preference for bold fonts (both serif and sans-serif) at close distances with large text. At comfortable or close distances (0.5 m–5 m), serif fonts with reduced weight improve legibility for complex Chinese characters (see panels 1, 2, 3, and 5 in Figure 5.8), whereas light fonts such as SaT and SeE remain unsuitable for simple glyphs (see panels 3 and 6 in Figure 5.8).

An additional consideration in the Japanese experiment is the role of kana. Katakana (J2), characterized by predominantly straight strokes, and Hiragana(J1), known for its curvilinear features, exhibit different font preferences. At closer distances (0.5 m and 2.5 m), Hiragana tends to be more suitable for serif fonts, regardless of weight. However, at farther distances (5 m and 10 m), the differences between the two become less pronounced.

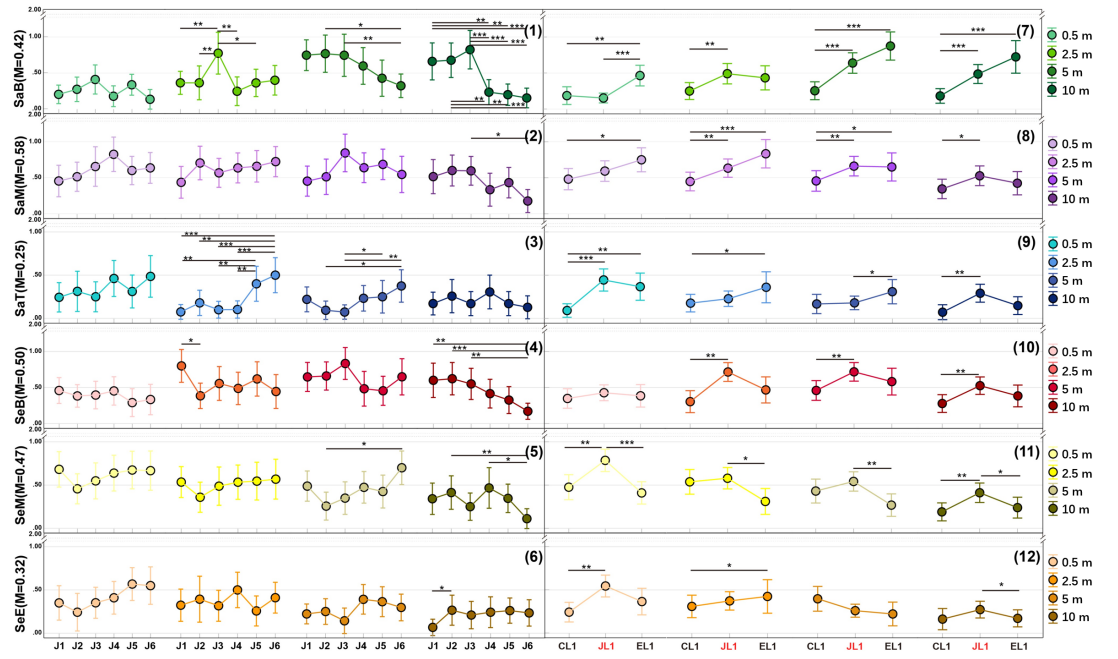



Fig. 5.8: Font Performance Trends of Japanese Fonts Under Varying Glyph Complexity, Native Language Groups, and Viewing Distances. Panels (1)–(6) illustrate how font performance varies with different glyph complexities (x-axis), while Panels (7)–(12) show how performance differs across native language groups (x-axis). In all panels, the y-axis indicates the mean font performance scores (0–2 scale) with 95% confidence intervals, where "M" denotes the average score for each font. Separate post hoc analyses were conducted for each trend, with significant differences marked by "*" ($*p < 0.05, **p < 0.01, ***p < 0.001$).

In the English experiment, sans-serif fonts generally outperform serif fonts marginally. Complex glyph forms (i.e., complex words) did not significantly affect legibility, and detailed comparisons did not reveal clear trends (see panels 1–6 in Figure 5.9). At close distances (0.5 m), medium-weight fonts demonstrate optimal performance. As distance increases (2.5 m-5 m), both medium and bold weights show good results, with bold fonts (SaB, SeB, SeM) being strongly preferred by Japanese speakers (JL1). additionally, JL1 participants increasingly preferred bold fonts like SaB and SeB at 5 m and 10 m (see panels 7 and 10 in Figure 5.9), while native English speakers favored bold fonts at 10 m. Notably, light fonts (both serif and sans-serif) are unsuitable for VR presentation at 2.5 m-10 m.

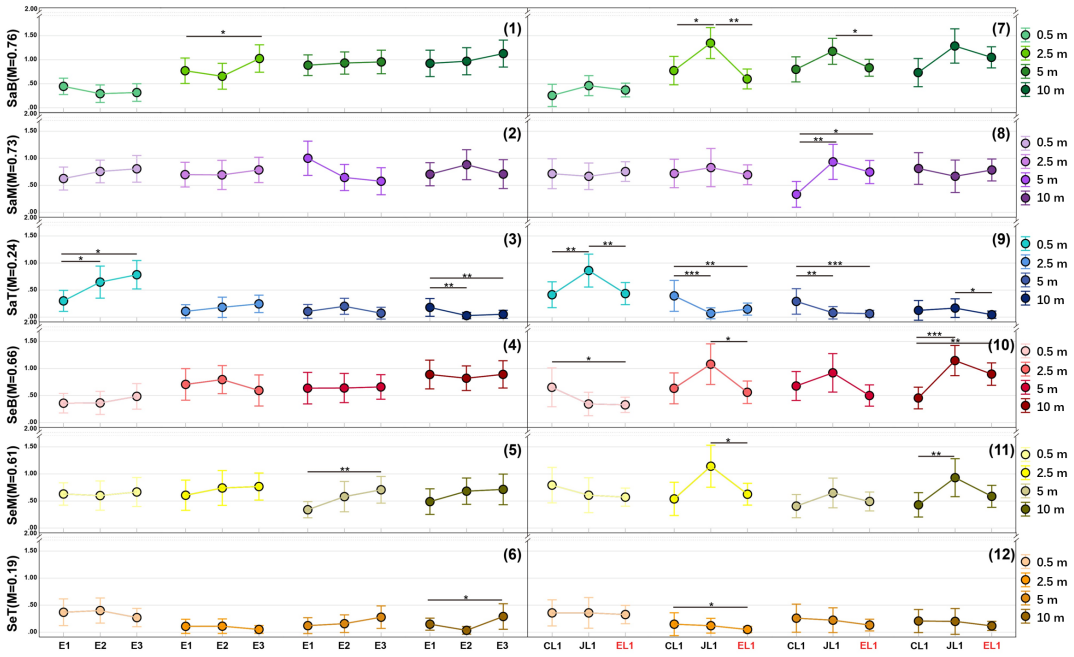



Fig. 5.9: Font Performance Trends of English Fonts Under Varying Glyph Complexity, Native Language Groups, and Viewing Distances. Panels (1)–(6) illustrate how font performance varies with different glyph complexities (x-axis), while Panels (7)–(12) show how performance differs across native language groups (x-axis). In all panels, the y-axis indicates the mean font performance scores (0–2 scale) with 95% confidence intervals, where "M" denotes the average score for each font. Separate post hoc analyses were conducted for each trend, with significant differences marked by "*" ($*p < 0.05, **p < 0.01, ***p < 0.001$).

## 5.4 Discussion of Key Findings

### 5.4.1 Distinct font preferences for Native and Non-Native speakers across language groups

Experimental results indicate that different native language groups exhibit distinct font preferences when processing both their native language and non-native scripts. Specifically, familiarity with a language leads to clear font preferences, whereas for unfamiliar languages, these preferences become less defined.

More precisely, the results demonstrate that, in each language-specific experiment, both native and non-native speakers showed varying font preferences between their L1 and non-L1 scripts. However, a detailed comparison in the Results section revealed that the influence of viewing distance on trends related to native language preferences was somewhat inconsistent. Although the Distance × Font × Native interaction was statistically significant across all language experiments (see Table 5.2), this result may

primarily reflect the stronger effects of the Font and Native factors. To clarify the language-related preferences, we excluded the distance factor and employed a heat map (Figure 5.10) to illustrate the trends in the Font × Native interactions across the different languages.

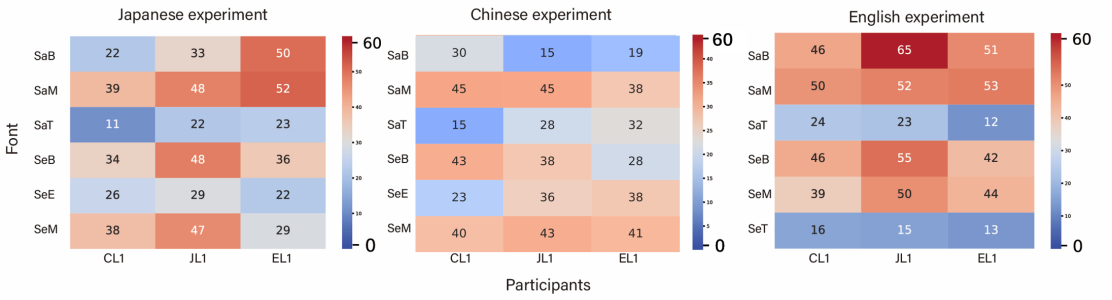



Fig. 5.10: Heatmap Illustrating Font × Native Interactions Across Different Language Experiments (Derived from Each Font' s Winning Rate)

To define regions of significant heat on the maps, a rule based on mean values and local differences was applied. First, the mean of all heat-map values was computed, and cells with values above this mean were preliminarily marked as high-value candidates. To more precisely identify areas of concentrated high values, any above-mean cell was further labeled as "concentrated heat" when its difference from surrounding above-mean cells was less than or equal to 10.

Figure 5.10 shows that, for Japanese fonts, native Chinese (CL1) and Japanese (JL1) speakers preferred Sans-MediumJP (SaM), Serif-MediumJP (SeM), and Serif-BlackJP (SeB) fonts, whereas native English speakers (EL1) favored Serif-BlackJP (SeB) and Sans-MediumJP (SaM) fonts. In the English experiment, both native (EL1) and non-native speakers (CL1 and JL1) exhibited similar preferences, focusing on Sans-Black (SaB), Sans-Medium (SaM), Serif-Black (SeB), and Serif-Medium (SeM) fonts, with a notable bias toward sans-serif typefaces.

For Chinese fonts, the heat maps revealed a more dispersed preference pattern, especially among non-native participants (JL1 and EL1). In contrast, native Chinese speakers (CL1) demonstrated a clear preference for Sans-MediumSC (SaM), Serif-BlackSC (SeB), and Serif-MediumSC (SeM) fonts, a pattern similar to their preferences for Japanese fonts. Conversely, non-native speakers (JL1 and EL1) showed more varied preferences for Chinese fonts, with a notably lower win rate for Sans-BlackSC (SaB).

This phenomenon suggests that when individuals are exposed to a familiar language (whether L1 or a well-practiced L2)—for example, when EL1 and CL1 are processing Japanese or when CL1 and JL1 are processing English—they tend to display similar and distinct font preferences, likely influenced by legibility considerations. In contrast, when the language is unfamiliar (L2), such as JL1 and EL1 encountering Chinese, font preferences are less pronounced. In these cases, aesthetic judgment may play a more

decisive role, particularly at closer distances where the letterform contours are clearly discernible. This finding supports earlier research by Blohm et al. [162]and Gao et al. [163], which suggests that fonts with lower inherent legibility can sometimes enhance aesthetic appreciation when semantic content is not the primary focus.

Detailed comparisons indicate that as the viewing distance increases—especially at the extreme limit—both native and non-native participants shift their criteria toward recognizability of letter contours . For complex Chinese and Japanese characters, neither bold nor thin typefaces significantly improve overall preference, implying that a larger font size may be necessary. In contrast, for simpler Chinese characters, kana, and English words, non-serif bold fonts tend to perform better at longer, particularly extreme, viewing distances.Everyday visual environments shape font biases across scripts.

Finally, although direct empirical data are limited, cultural exposure may further shape these biases. JL1 participants consistently favored bold weights across all language contexts—perhaps reflecting the ubiquity of heavy lettering in Japanese signage and advertising. Likewise, in English experiment, sans-serif and bold styles were strongly preferred by every group. Mackiewicz found that academic audiences rated sans-serif PowerPoint text nearly twice as readable as serif text [164], and because the sample in this study comprised mostly university students, habitual exposure to such styles may have influenced their legibility judgments.

### 5.4.2 Impact of Font Style and Weight on Preferences and Design Implications

Typefaces are primarily defined by their style (serif versus sans-serif) and weight (thin, medium, bold). Because medium-weight fonts consistently exhibited the best performance, the analyses in this section focus on trends between bold (black) and thin (light) fonts, with medium weights omitted for clarity. Figure 5.11 illustrates these trends, with blue lines representing bold fonts and orange lines representing thin fonts. Dark-colored lines denote serif fonts, while light-colored lines denote sans-serif fonts. The x-axis indicates the typeface design (i.e., glyph complexity), and the y-axis shows the win rate (calculated as the number of wins divided by the total number of appearances) for each font across different glyph complexities.

Clear trends emerged across languages. In both the Japanese and Chinese experiments, a crossover pattern appeared as glyph complexity increased: bold fonts tended to show a decrease in win rate, whereas thin fonts exhibited an increase. This suggests that, for more complex characters, bold fonts become less favorable while thin fonts become more favorable. In contrast, English showed a more parallel pattern, indicating that glyph complexity had minimal impact on font performance; bold English fonts consistently achieved higher win rates than their thin counterparts. This pattern is consistent with the overall results, in which English font preferences were more strongly influenced by weight, whereas Japanese and Chinese preferences were more sensitive to

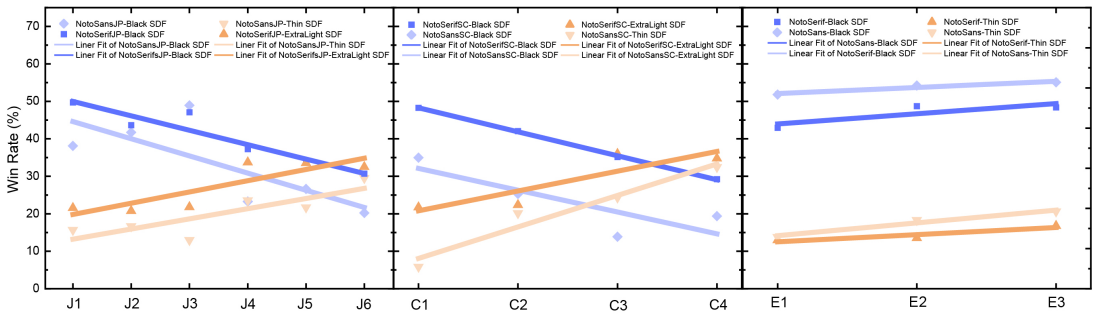



Fig. 5.11: Win Rate Trends of Fonts Across Languages and Glyph Complexity Levels

glyph complexity. These findings also align with prior work (e.g., Oderkerk et al. [131]), which showed that increasing letter width improves letter recognition—particularly for shorter and wider shapes—while the complexity of Chinese and Japanese characters modulates font preferences differently: simpler characters tend to favor bolder weights, whereas more complex characters benefit from thinner strokes.

Regardless of complexity, serif fonts consistently outperformed sans-serif fonts for both Chinese and Japanese . In contrast, for English, sans-serif fonts slightly outperformed serif fonts. This reinforces the notion that serif fonts are generally preferred for more complex scripts like Chinese and Japanese—especially in close-range contexts—whereas sans-serif fonts are more suitable for English, particularly at longer viewing distances.

In summary, the findings indicate that font weight and style preferences are systematically shaped by both the inherent complexity of the script and the language context. Bold fonts are typically favored for simpler characters, whereas thinner fonts are preferred for more complex characters, especially in Japanese and Chinese. These patterns have important implications for typography and interface design across different languages.

### 5.4.3 Summary of Findings

While initial GLM analyses revealed multiple factors significantly influencing font preferences in the Chinese (Section 5.3.2), Japanese (Section 5.3.2), and English (Section 5.3.2) experiments, a detailed examination of the results in Section 5.3.2 and subsequent discussions identified two primary determinants: the interaction between viewing distance and glyph complexity (Font × Distance × Glyph), and native language backgrounds (Font × Native). Figure 5.12 synthesizes these findings to visualize font preference patterns across diverse conditions.

First, for Chinese and Japanese, the interaction between glyph complexity and viewing distance had the greatest impact on font preference. To illustrate varying levels of glyph complexity, representative characters were selected at each level (e.g., J1–J6 for Japanese

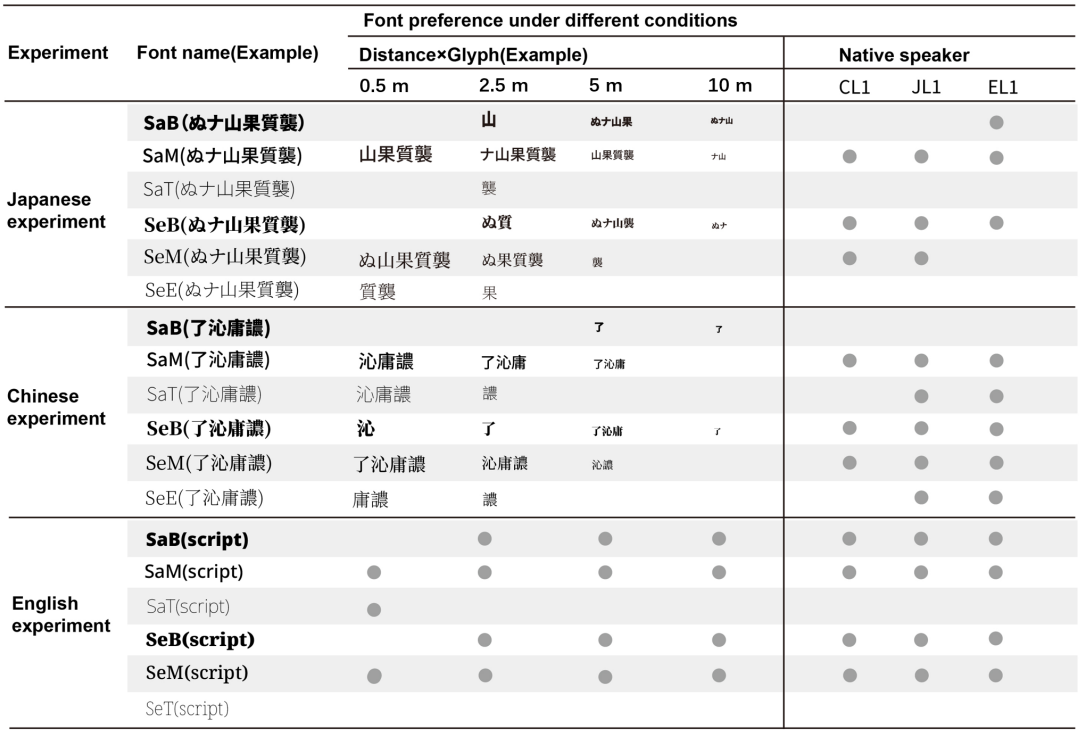



Fig. 5.12: Comprehensive Overview of Font Preferences by Distance × Glyph, and Native Speaker

and C1–C4 for Chinese). Because English font preference did not significantly depend on glyph complexity, the word "script" (E2) was used as a proxy item. Drawing on the detailed comparisons reported in the Results section, fonts that performed relatively well at specific distances and for different glyphs were identified (acknowledging that VR experiences differ substantially from screen-based viewing, but using font size as an approximate analogue of distance). An average-score threshold of 0.5 was used to judge whether a font performed well. For English, glyph complexity showed no significant effect, whereas the Distance × Font interaction did (see Table 5.2); accordingly, a dot marker is used in Figure 5.12 to denote these cases.

Second, although the initial ANOVA indicated that native language, Native × Font, and Native × Font × Distance were all significant, more detailed analyses did not reveal a stable pattern for the three-way interaction. The discussion therefore focuses on the Native × Font level. In Figure 5.12, the "Native" portion is primarily based on the overall preferences shown in Figure 5.10, with specific font examples added to clarify the observed patterns.

The key findings can be summarized as follows:

1. **Different font weights showed distinct preferences across varying distances.** In Chinese and Japanese, heavier fonts and simpler glyphs were preferred

at greater distances. At closer distances (0.5 m, 2.5 m), medium-weight fonts generally performed well across most glyphs, while lighter fonts were more suitable for complex glyphs. In contrast, English results indicated that thinner fonts consistently performed poorly, and heavier fonts were less preferred at the closest distance (0.5 m). Therefore, in VR environments, it is advisable to avoid thin fonts for English and to refrain from using heavy fonts at 0.5 m, regardless of whether they are serif or sans serif.

2. **In the Chinese and Japanese experiments, serif fonts outperformed sans-serif fonts at non-extreme distances (0.5 m–5 m).** At closer distances—especially 0.5 m—serif fonts provided a clear advantage because their decorative details remained clearly visible, particularly for complex glyphs (J5, J6; C3, C4). However, bold serif fonts (SeB) were unsuitable at the very closest distance. Within Japanese scripts, Hiragana tended to benefit more from serif fonts than Katakana. A plausible explanation is that Katakana (e.g., ナ, フ), composed mostly of straight strokes, resembles simpler Chinese characters (e.g., 一, 七) and therefore gains relatively little from additional decorative elements, whereas Hiragana (e.g., な, ふ), with its more curved strokes, appears more decorative and thus benefits more from serif detailing.

3. **Familiarity with glyphs leads to more consistent and pronounced preferences.** For example, because English is widely recognized as an international language, CL1 and JL1 participants showed nearly identical font preferences for English. In the Japanese experiments, CL1 participants— familiar with kanji and often learning Japanese as a second language (regardless of proficiency)—exhibited preferences similar to JL1 participants. In contrast, EL1 participants, whose native language is alphabetic, displayed different preferences. Finally, in the Chinese experiments, both EL1 and JL1 participants, with no prior study of Chinese, showed more variable preferences.

4. **When participants encountered entirely unfamiliar glyphs, their preference criterion shifted from "legibility" to "visual appeal."** Although the task instructions emphasized legibility, the semi-structured interviews revealed that complex or unknown characters (e.g., certain Chinese glyphs) were primarily evaluated based on perceived attractiveness. Consequently, the second finding—that serif fonts are preferable at close distances—should be interpreted with caution. As viewing distance increases and decorative details become less discernible, however, participants across language backgrounds tend to favor whichever font appears clearer.

## 5.5 Chapter Summary

The study reported in this chapter conducted a cross-linguistic VR experiment to systematically investigate font preferences among native and non-native speakers across viewing distances (0.5 m–10 m) and glyph-complexity levels. Participants performed pairwise font comparisons in three language contexts—Chinese, Japanese, and English—within an immersive virtual environment, generating distinct preference profiles. Through quantitative analysis of interaction effects (Font × Distance × Glyph × Native) and qualitative interviews, language-specific adaptation patterns were identified and translated into evidence-based design guidelines.

Although the study is limited by the lack of semantic context and a restricted font set, it nevertheless offers three main contributions to VR text adaptation. First, it provides empirical evidence of cross-linguistic divergence in font preferences, showing that Chinese characters in Chinese and Japanese contexts benefit from serif fonts at close distances (0.5 m–2.5 m), while requiring larger weight adjustments for complex glyphs than alphabetic systems (English), consistent with prior findings on orthographic processing [131]. Second, the results highlight the dynamic interplay between aesthetic preferences and functional legibility, particularly for non-native readers who often prioritize visual appeal over legibility in unfamiliar scripts, in line with dual-process accounts of typographic perception [162, 163]. Third, the chapter derives practical design guidelines:

- For English interfaces across varying distances, prioritize sans-serif fonts and increase font weight as distance increases;
- For Chinese and Japanese texts at close distances, adjust font weight according to glyph complexity (using lighter weights for complex characters and heavier weights for simpler ones);
- At extreme viewing distances, rely primarily on adaptive text sizing rather than weight adjustments alone.

By integrating psycholinguistic principles with VR-based experimentation, this work contributes a methodological framework for evaluating perceptual–cognitive interactions in immersive environments and proposes concrete strategies to enhance digital inclusivity. The findings suggest that VR interfaces can better accommodate diverse linguistic populations through empirically grounded font choices. Future extensions that incorporate semantic content and expand the range of typographic variables may further refine these guidelines and advance VR as a medium for multilingual communication.

# Chapter 6 Discussion

This chapter brings together the findings from the three experimental studies by adopting an MTMM method: the preferred viewing distance task, the glyph identification task, and the subjective font rating task are treated as complementary methods probing shared traits of VR text legibility and preference, as well as task- and language-specific effects. In the first part of the discussion, the empirical patterns are translated into more general design principles and the fine-grained guidelines are consolidated into an initial set of rules for a real-time legibility support system (VTLS), as a preliminary attempt to operationalize the results in practice. The subsequent part examines differences between native and non-native readers, focusing on how perceptual and cognitive requirements shift when users read in a non-native language and outlining early design ideas that might better support second-language reading in VR. The chapter concludes by reflecting on the limitations of the present work and by indicating several directions for future research.

## 6.1 Operationalization of VR text-legibility findings

### 6.1.1 Language-Independent VR Text Legibility Guidelines

**1) Thin Latin fonts are unsuitable for VR** Converging evidence across all three studies supports the view that thin (light) English fonts are ill-suited to VR displays. In Chapter 3, thin Latin strokes required shorter viewing distances at every tested range (0.5 m–10 m; see Fig. 2.3b), consistent with a pixel-threshold account: as distance increases, high-frequency stroke details are undersampled and edge energy collapses. In Chapter 4, at 5 m the medium and thin serif settings (SeM, SeT) slightly increased errors for the E3 set. Chapter 5 further showed that light weights underperform across distances, with especially poor clarity at far viewing (2.5 m–10 m) where thin strokes vanish first. Therefore, prefer medium or bold weights for English in VR to preserve boundary contrast as distance grows.

**2) At near/comfortable distances, lighter weights can improve legibility for complex glyphs** For highly folded, high-stroke-density forms (complex Kanji/Hanzi), Chapter 5 found that at near or comfortable distances (0.5 m–5 m) lighter weights often improved perceived clarity by reducing within-character crowding and making fine structure easier to parse. This is a distance-bound effect and does not generalize to far viewing.

**3) Simple glyphs paired with heavier weights read farther** For low-complexity glyphs,whether simple Hanzi/Kanji, simple kana, or Latin letter, heavier weights enhance distance legibility by strengthening stroke boundaries under reduced sampling. Chapter 3 showed that simple Kanji/Hanzi and kana in bold remained recognizable at 10 m more than lighter counterparts. Chapter 5 replicated this in Chinese and Japanese and noted that in English interfaces the stroke weight should be increased as viewing distance grows.

**4) For complex glyphs at far distance, increase size or reduce distance** Font tuning alone is insufficient at 10 m for complex characters. In Chapter 3, complex Hanzi at 10 m were nearly indiscriminable even for L1 readers, indicating that subjective visual compensation fails under high stroke density; moving the panel closer was necessary. Chapter 4 quantified that higher complexity raises error and invisibility rates at far distances, especially for non-natives. Chapter 5 observed that although thin font can locally raise preference for complex forms, absolute scores at 10 m remain low. Therefore, for highly folded, complex glyphs, the only robust options are to enlarge the text or shorten the viewing distance

**5) At near distance, hiragana benefits from serif faces** Chapter 5 showed that at 0.5 m (and often 2.5 m) hiragana tended to be judged more acceptable in serif faces, with this preference especially strong among Japanese native readers. A plausible account is that hiragana' s curvilinear strokes can exploit resolvable serif detailing at short distances, reinforcing contour cues.

**6) English shows no reliable serif advantage; sans-serif is clearer at distance** Across English conditions, no consistent preference for serif was observed. Chapter 5 reported a small overall sans-serif advantage that becomes more apparent as distance increases, reflecting simpler letterforms whose clarity depends more on stroke weight than on serifing. Combined with Chapter 3' s evidence that thin strokes are fragile at distance, the practical rule is to avoid light weights; medium/bold weights are safer as distance increases, and the serif–sans choice is secondary to weight and viewing distance in VR.

### 6.1.2 Initial attempts at future work: VTLS- A real-time legibility system for VR text based on VR guidelines

This section describes how the empirical findings from Chapters 3–5 were translated into a run-time mechanism that automatically adapts text size and font style to the viewer' s position in virtual space. The goal is to bridge experimental, language-independent legibility guidelines and an implementable module that can be embedded in VR/AR applications. The proposed VR Text Legibility System (VTLS) (i) takes as

input the current camera–text distance and the linguistic/glyph attributes of the text, (ii) computes an appropriate world-space text height through a distance-sensitive scaling rule, and (iii) selects a font/weight configuration that conforms to the guidelines derived from the user studies. The design is intentionally modular so that future guidelines (e.g., on spacing or colour contrast) can be incorporated with minimal changes to the core scaling rule.

### Algorithm

The run-time adaptation builds on two ingredients: the language-independent VR legibility guidelines established in Chapters 3–5, and the distance-specific predictions produced by the LME/EMM analysis of Experiment 1. The guidelines define what should happen qualitatively (e.g. thin Latin forms deteriorate first at distance; complex CJK characters require enlargement or reduced distance; kana can be kept relatively small at comfortable ranges), while the EMM tables provide concrete distance offsets for each language–glyph combination. Experiment 1, however, sampled four distances, $D_1 = 0.5\,\mathrm{m}$, $D_2 = 2.5\,\mathrm{m}$, $D_3 = 5\,\mathrm{m}$, and $D_4 = 10\,\mathrm{m}$, where $D_2$ and $D_3$ correspond to "comfortable near" and "comfortable far", and $D_1/D_4$ were included to probe the lower and upper visibility limits. In a real-time system it is still useful to make use of these two extreme points: very near text should not grow indefinitely, and very far text (10 m and beyond) should still receive a compensatory enlargement. The formulation below therefore keeps $D_1$ and $D_4$, but uses them as smoothly tapering bounds rather than as equally strong targets.

Let $d$ denote the Euclidean distance (metres) from the user' s camera to the text anchor at run time. Let $L$ be the language, $c$ the glyph/complexity class, and $d^*_{L,c}$ the preferred distance estimated from the EMM analysis for that $(L, c)$ pair (typically close to $D_2$ or $D_3$). Let $h_0$ be the baseline world-space text height ($\approx 0.07\,\mathrm{m}$ in the prototype), and let $h_{\min}$ and $h_{\max}$ be the application-level bounds.

To retain the fine-grained differences between simple and complex characters, each script is assigned a normalised complexity value $\mathrm{lever}_{L,c} \in [0, 1]$ derived from the EMM values at $D_2$ (and, when needed, at $D_3$): for Chinese, this produces an increasing staircase $C1 \to C2 \to C3 \to C4$; for Japanese, it produces $J3$ (easy Kanji) $\approx 0$, hiragana and katakana slightly above 0, and $J4 \to J5 \to J6$ as a rising tail for complex Kanji. This lever does not replace the guidelines; it only makes the size change progress in measurable, non-collapsing steps.

Distances shorter than $D_2$ and longer than $D_3$ are shaped by two simple factors. A near factor

$$s_{\mathrm{near}}(d) = \begin{cases} \dfrac{d - D_1}{D_2 - D_1}, & D_1 \le d < D_2, \\ 1, & d \ge D_2, \end{cases}$$

prevents very close text ($\approx 0.5\,\mathrm{m}$) from overreacting, while a far factor

$$s_{\mathrm{far}}(d) = \begin{cases} 1, & d \le D_3, \\ 1 - \dfrac{d - D_3}{D_4 - D_3}, & D_3 < d < D_4, \\ 0, & d \ge D_4, \end{cases}$$

ensures that the EMM-based complexity correction gradually fades out between 5 m and 10 m, after which only the main distance scaling remains. For distances beyond 10 m ($d \ge D_4$), the system keeps enlarging the text through the distance term, but the script-specific lever is no longer amplified, so extremely far panels do not become disproportionately large.

The run-time height is then defined as

$$h(d; L, c) = \min\Big\{\max\big\{h_0\big[1+\beta\big(\frac{d}{d^*_{L,c}} - 1\big)\big]\,\big(1+\alpha\,\mathrm{lever}_{L,c}\,s_{\mathrm{near}}(d)\,s_{\mathrm{far}}(d)\big),\ h_{\min}\big\},\ h_{\max}\Big\},$$

where:

$\beta \in (0,1)$ attenuates sensitivity to small distance changes, so that head jitter does not trigger visible size jumps;

$\alpha$ controls the influence of the EMM-derived complexity term (values in the range 0.2–0.4 keep the modulation gentle);

$s_{\mathrm{near}}(d)$ introduces the information from $D_1$ without forcing extreme enlargement at very close ranges;

$s_{\mathrm{far}}(d)$ introduces the information from $D_4 = 10\,\mathrm{m}$ but lets it decay smoothly, so that panels slightly beyond 10 m are still enlarged by the distance term, not dominated by per-script corrections.

In this expression, the bracket

$$h_0\big[1 + \beta\big(\tfrac{d}{d^*_{L,c}} - 1\big)\big]$$

is the language- and complexity-aware distance scaling: if the current distance matches the preferred distance ($d = d^*_{L,c}$), the text is rendered at the baseline height; if the user stands farther away, the height grows in proportion to $\frac{d}{d^*_{L,c}}$ but moderated by $\beta$; if the user stands nearer, the height is reduced but never below $h_{\min}$. The multiplicative term

$$1 + \alpha\,\mathrm{lever}_{L,c}\,s_{\mathrm{near}}(d)\,s_{\mathrm{far}}(d)$$

injects exactly the information that the EMM revealed: within the comfortable window $[D_2, D_3]$, complex Chinese and Japanese characters receive slightly larger text than simple ones; at very near distances this correction is softened; at very far distances it is




Fig. 6.1: Base prototype with neutral blocks and placeholder text.

gradually switched off and only the global distance scaling remains. The outer min–max pair preserves layout-level constraints.

After the height is determined, font family and stroke weight are assigned in accordance with the guidelines: thin Latin is avoided when $d \geq d^*_{L,c}$; complex CJK at $d > d^*_{L,c}$ is rendered in a sans-serif bold face; near hiragana may keep a serif/Mincho face. Keeping the style decision on top of (rather than inside) the height equation allows later work on spacing or colour contrast to reuse the same numeric core.

In practice this means that, when the user stops moving, all text panels currently visible in front of the camera are rescaled so that (i) close panels do not become oversized, (ii) mid-range panels follow the exact language–complexity preference measured in VR, and (iii) very distant panels, including those at 10 m and slightly beyond, are still brought into a readable size instead of remaining unchanged.

### Prototype implementation

As a first step towards a concrete system, a simple prototype scene was created in Unity using only neutral blocks and planar panels. Each block carries a world-space canvas with placeholder text; the content itself has no semantic meaning and all panels share the same initial font and size. The scene therefore serves as a controlled base in which only distance and orientation vary. At discrete moments—when the headset motion falls below a small velocity threshold—the system updates each visible text object according to the VTLS-style rules: the distance $d$ and the language/complexity category $(L, c)$ determine a target angular size, and font weight and scale are adjusted accordingly. The aim here is not to implement the full formal model with high numerical precision, but to visualise how distance-aware and script-aware scaling would behave in an interactive environment. Figure 6.1 shows three views from this prototype, illustrating how text on nearer blocks becomes relatively larger and heavier, while distant panels remain readable without manual resizing.

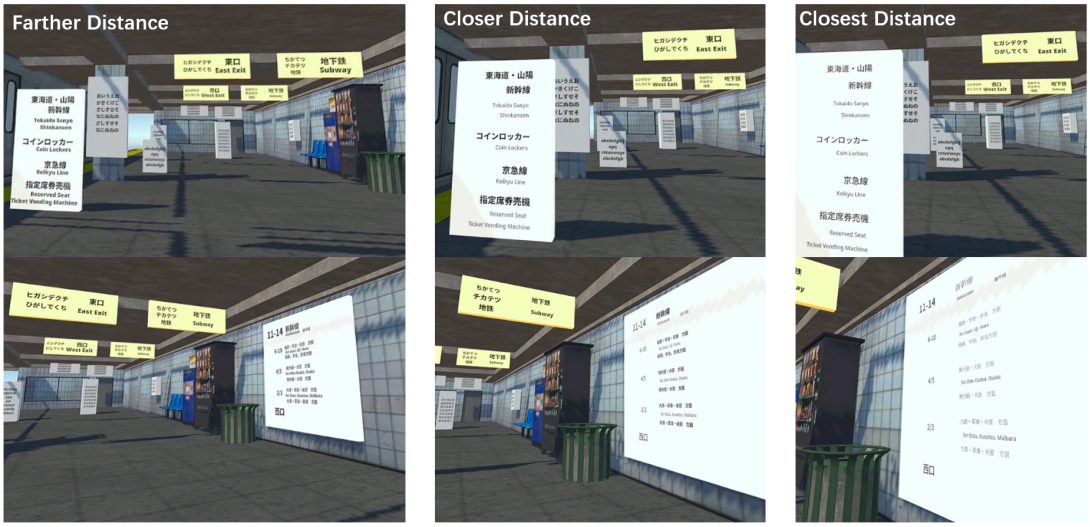



Fig. 6.2: Metro-style scene that augments the base prototype with simple information-hierarchy rules on top of distance- and complexity-aware text adaptation.

**Design-rule augmentation: metro-signage example**

Building on this base, a second scene extends the prototype towards a metro-signage scenario. Here the same distance- and complexity-based adjustments are combined with simple design rules that reflect how information is typically prioritised in transport hubs: station names and exit directions are rendered larger than secondary notes, warning panels are forced into high-contrast sans-serif styles, and less critical details are slightly down-weighted in size. The resulting station-like environment (Figure 6.2) is not intended as a finished design, but as an initial exploration of how experimentally grounded legibility thresholds can be layered onto familiar hierarchies such as "H1/H2 larger, body text smaller" while keeping the text responsive to viewer position.

Because the legibility decisions are available in symbolic form ($(L, c)$, $d$, $d^*_{L,c}$, chosen font and weight), these additional rules can be written declaratively, without re-deriving the perceptual model. At the same time, important questions remain open: how strongly different information categories should be separated, how to coordinate multiple scripts on the same sign, and how users with different language backgrounds experience the resulting layouts. Addressing these issues, and validating such VTLS-augmented signage with real travellers and tasks, constitutes a natural direction for future applied work.

## 6.2 Perceptual–Cognitive Transitions between Native and Non-Native Readers in VR

### 6.2.1 Visual Requirements

This work used four viewing distances, with 0.5 m and 10 m as extremes and 2.5 m–5 m as comfortable ranges (0.5 m–2.5 m = near; 5 m–10 m = far).

When text size and resolution already satisfy basic perceptual needs, native and non-native readers show divergent near-distance preferences, especially for font choices. In Chapter 5, clear divergences appeared at 0.5 m–2.5 m:

- English task: At 2.5 m, JL1 tended to prefer SaB, SeB, and SaM; CL1 preferred SeB at 0.5 m but shifted to SaT at 2.5 m.
- Japanese task: JL1 preferred hiragana with SeB, whereas EL1 leaned toward SaB and SaM and, at 2.5 m, toward SaM/SaT/SeE for complex glyphs (thin weights being mainly suitable for complex forms at near/comfortable distances).

As distance increases, perception and decision-making shift toward a clarity-first criterion, and the above preferences converge. For example, In Chapter 5, for Japanese/Chinese, simple glyphs pair best with heavier weights at distance, while complex glyphs cannot be improved by font choice alone; for English, thin fonts are not suitable at 2.5 m–10 m. Consistent with this pattern, Chapter 3 showed that at 10 m all groups agreed that "simple + bold" extends recognizable range, whereas complex forms must be brought closer (or enlarged). Therefore, With greater distance and coarser sampling, processing is reweighted from perceptual, familiarity-based cues (Type 1) toward analytic, cognition-heavy strategies (Type 2), with legibility taking priority.

### 6.2.2 Role of Experience

Experience (as L1 or as a well-practiced L2) tends to produce more consistent judgments; in contrast, little or no exposure leads to scattered or aesthetic-driven responses. In Chapter 5, given that participants were recruited in Japan, CL1 generally had some baseline exposure to Japanese and English, JL1 had baseline English exposure, and EL1 had baseline Japanese exposure. Consequently, experienced L2 readers produced clearer non-L1 preferences, whereas readers without such exposure showed less stable patterns (see Fig. 4.10).

Experience also supports perception-led shortcuts when stimuli are challenging. In native-script tasks, perceptual and cognitive processes run in parallel, and leaning on perception reduces cognitive load. For instance, in Chapter 3 at far distance, JL1 and EL1 exhibited visual compensation relative to non-natives, enabling recognition at greater distances. In Chapter 4, as glyph complexity increased in Japanese and Chinese, CL1/JL1 did not show rising error rates or longer completion times, whereas EL1

errors rose sharply and responses slowed. A plausible account is that experience enables contour/radical-level parsing (perception-led), while inexperienced readers must rely more on analytic checking (cognition-led). In short, compared with non-natives, experience amplifies perceptual processing and reduces cognitive load via top-down support.

### 6.2.3 Influence of Cultural Exposure

Native visual recognition habits shape how readers process non-native scripts in VR. Readers bring over the routines they have practiced for years, and those routines steer both perception (fast, familiarity-based) and cognition (slower, analytic) when the script changes.

In Chapter 4, logographic L1 readers (CL1/JL1) relied on contour-based, familiarity-driven processing that supports speed, but this same habit made them more vulnerable to similarity-based confusions. When switching to English, they often treated words "as shapes," which preserved pace yet reduced accuracy for confusable pairs; by contrast, EL1 readers expended comparable time but kept errors low by using alphabetic, part-based checks.

Chapter 4 also showed the mirror image for EL1 on Hanzi/Kanji: alphabetic routines did not transfer well to highly folded, multi-stroke characters. EL1 maintained low similarity errors but read more slowly, and both errors and completion times rose as complexity increased—evidence of a shift toward effortful, analytic processing when familiar perceptual cues fail.

Chapter 3 further illustrated that L1-tuned spatial sampling preferences persist even in non-native contexts. At near distances, EL1 placed the panel slightly farther to reach a comfortable limit, consistent with a larger, right-asymmetric perceptual span and the benefits of parafoveal preview in alphabetic reading. Together, these findings indicate that cultural exposure calibrates which visual cues are treated as diagnostic, and these priors continue to guide processing in L2; as distance increases and effective resolution drops, those priors are reweighted toward clarity-first decisions.

### 6.2.4 Initial attempts at future work: Design Implications for Native and Non-Native Readers

Finally, this dissertation proposes a few exploratory directions. These brief notes add no new claims; they simply outline testable ideas—as prompts for future prototypes.

#### Helping Alphabet Natives Recognize Japanese/Chinese

Alphabet L1 readers (e.g., EL1) rely more on analytic checking when facing many-stroke characters and are sensitive to complexity and similarity (Chapter 4). The fol-

lowing practices reduce cognitive load and similarity confusions:

1. **Group strokes into perceptual blocks (radicals/components).** Prefer renderings that expose component boundaries (e.g., outline consolidation, local whitespace, subtle haloing) so that complex characters are perceived as 2–3 chunked "units." This lowers near-distance (0.5 m–2.5 m) similarity interference and yields more stable outer-shape cues at 5 m–10 m (Chapter 4/5).

2. **Use consistent color coding as an optional training layer.** Apply a stable mapping between colors and semantic/phonetic components during learning or hint states; fade the color layer at 5 m–10 m to protect global contrast and stroke continuity, in line with the clarity-first pattern from Chapter 3.

3. **Avoid overly decorative styles; prefer medium–bold weights.** Effects that break contour continuity (glow, extreme contrast, heavy shadows) harm component parsing. From 2.5 m–10 m, medium or bold Gothic/Hei styles prevent hairlines from dropping out; thin styles are not suitable for English and offer little help for complex Kanji/Hanzi at distance (Chapter 3, Chapter 5).

4. **Offer "closer/larger" as the primary aid for high-complexity glyphs.** When complexity exceeds a threshold (e.g., $C > 2$), prioritize gaze-triggered zoom, hover-to-highlight components, and side-by-side exemplars over font swapping, which matches the similarity-error pattern in Chapter 4.

### Leveraging Logographic Natives' Perceptual Habits to Support English

Logographic L1 readers (CL1/JL1) tend to treat English words as shapes: recognition is fast but confusable pairs can degrade accuracy (Chapter 4). Rather than emphasizing micro-letter differences, a block-first approach aligns better with their visual routines:

1. **Chunk words into stable blocks (prefix–root–suffix) and make the blocks the primary units.** Present unfamiliar words as 2–3 visual modules with subtle backgrounds, hairline brackets, or spacing cues to delineate prefix–stem–suffix. Maintain a consistent block palette across the UI so the whole word becomes a stable "shape of blocks". This preserves speed and reduces shape-based confusions without relying on fine letter diagnostics (consistent with near-distance font-preference divergences in Chapter 5).

2. **Guide input and correction by blocks, not by single letters.** Provide block "slots" for typing; enable block-level autocomplete and spell-checking (e.g., confirm *un*–**believe**–*able* in three steps). When errors occur, highlight the incorrect block and suggest replacements at the morpheme level, minimizing interruption while leveraging whole-shape familiarity.

3. **Choose type that preserves stroke continuity at distance.** At 5 m–10 m, avoid hairline and high-contrast thin serifs; prefer medium weights in sans or robust serifs to keep ascenders/descenders and counters open, in line with the clarity-first pattern (Chapter 3/5).

4. **Use lightweight scaffolds that decay with distance.** Allow temporary block segmentation and morphology hints at 0.5 m–2.5 m; gradually retire these as distance increases so global contrast and legibility dominate (Chapter 3).

**Distance-adaptive rendering (cross-group).** A practical rule set unifying Chapters 3–5 is: R1) Pair simple glyphs with **bold** weights at 5 m–10 m; R2) For complex glyphs, prioritize **closer/larger** over font changes; R3) Allow **training layers** (color coding, block segmentation) at 0.5 m–2.5 m, then **fade them** as distance grows to protect contrast and stroke continuity; R4) Maintain configurable **confusability lists** (English letter/word pairs; Kanji/Hanzi look-alikes) to trigger block-level hints or zoom on demand.

These recommendations couple experience-driven "perceptual shortcuts" with distance-dependent sampling limits: near distances retain familiar cues to reduce cognitive load, while far distances shift decisively to structure simplification and weight reinforcement to secure robust legibility across native backgrounds.

## 6.3 Limitations and Directions for Future Research

1. **Shared cohort and setup.** All three studies used the same participant cohort, the same HMD (Quest 3), the same Google Noto font family, and the same viewing-distance set. This design strengthened internal consistency and made cross-study comparisons straightforward, but it also narrows generalizability and introduces the possibility of carryover (practice, fatigue, strategy transfer). Future studies could separate testing into distinct waves or schedule repeated participation at different times, which would reduce dependency and make the findings more broadly applicable.

2. **Typography scope.** Fonts varied only within Noto (serif/sans; thin/medium/bold). The intent was to employ familiar, widely legible faces, which supports representativeness; however, the relatively modest stylistic spread likely dampened typographic effects (as seen in Studies 1 and 2). In future tests—especially those focused on typographic effects—it would be useful to include designs with stronger stylistic range (e.g., calligraphic or display/advertising styles alongside conventional text faces) to probe whether larger differences yield clearer patterns.

3. **English case restriction.** English materials were predominantly lowercase. Effects tied to case structure (all caps, Title Case, mixed case) were not assessed and

may influence legibility. Including these conditions as part of the glyph-variation space would provide a more complete picture.

4. **Task context.** The tasks emphasized single characters and words, or preference/threshold judgments. Sentence- and paragraph-level comprehension, semantics, and layout factors were outside scope, which limits ecological validity. Building on the present findings, future work can extend text length and explicitly control page/VR layout so that design implications are tested under more realistic reading conditions.

5. **Sampling and power.** Recruitment in Japan led to uneven L2 distributions. Many non-Japanese participants naturally used Japanese as an L2, whereas exposure to Chinese was often limited, producing imbalances by language background. Future work can recruit globally or to run multi-site studies that separately sample L1 and L2 groups across locations, improving balance and coverage.

6. **Age and vulnerable populations.** Participants were visually healthy younger adults, and within this cohort the main systematic difference was native language background. As such, the present findings should not be generalized to older adults, low-vision readers, or other vulnerable populations. Nevertheless, the basic approach—controlling VR conditions and systematically manipulating visual factors (e.g., glyph complexity, font weight, viewing distance) to compare groups —can be applied in future work that explicitly samples these populations.

7. **Unresolved "Chinese compensation."** Unlike English and Japanese natives, native Chinese readers did not show a clear far-distance compensation advantage here. Chapter 3 discussed possible reasons, but the question remains open and merits exploration from additional angles.

8. **Device and physiology controls.** All data came from a single HMD in a controlled lab. Interpupillary distance was adjusted for comfort but not recorded numerically, and visual acuity was self-reported rather than screened; both choices can add unmodeled variance to legibility outcomes. Future studies may record IPD and conduct brief vision screening so that these factors can be modeled rather than treated as background noise.

# Chapter 7 Conclusion

This dissertation set out to explain how visual factors—viewing distance, glyph complexity and similarity, and elementary typeface features—shape text legibility in VR for native (L1) and non-native (L2) readers of Chinese, Japanese, and English. Three complementary studies were designed around a shared parameter set (same participant cohort, headset, font family, sizes, and distances) so that perception, rather than language comprehension, would be foregrounded. Chapter 3 investigated preferred viewing distance, Chapter 4 isolated glyph-level legibility under complexity and similarity, and Chapter 5 examined font preferences across scripts. Taken together, this converging-operations structure allows results from distinct tasks to be interpreted as facets of a common construct: how visual evidence is sampled and used for recognition in head-mounted displays.

Across these studies, several consistent patterns emerged. As viewing distance increases, decision criteria shift toward clarity, and recognition becomes less tolerant of fine detail: simple forms paired with heavier strokes remain legible farther away, whereas complex characters and thin strokes require closer placement. At near ranges (0.5–2.5 m), native-experience divergences become visible—for example, English natives are comfortable reading slightly farther from the panel—whereas at far range (10 m) familiarity extends usable distance for Japanese and English but not reliably for Chinese under the present design. The glyph-identification experiment clarifies why: structural familiarity accelerates processing for Chinese and Japanese natives yet heightens sensitivity to similarity, while English natives facing many-stroke non-native forms are limited chiefly by complexity, exhibiting longer completion times but being less driven by similarity. Typeface results align with this distance-driven account. Within the controlled Noto family, thin Latin styles become fragile beyond D2, and heavier weights improve robustness as distance grows. For highly folded Kanji and Hanzi at near or comfortable ranges, lighter weights can reduce within-character crowding, but at far distances font tuning alone is insufficient—content must be brought closer or enlarged. Hiragana often benefits from serif detailing at D1 (and sometimes D2), whereas English shows no reliable serif advantage; at distance, a small sans-serif edge emerges and stroke weight dominates style.

These integrated findings distill into practical, perception-first guidance for VR text: interfaces should be both distance-aware and script-aware, adjusting weight, size, and placement in line with language-specific constraints and reader background. Building on this, the dissertation proposed an initial Visual Text Legibility System (VTLS) that

encodes these guidelines into simple rules for real-time text adaptation, and outlined preliminary design ideas to better support non-native readers when working with second-language content in VR. Overall, the work clarifies how distance, glyph structure, and typeface interact with language background to shape reading in immersive displays, and offers a compact set of principles together with a first system concept. Future research can extend these ideas to richer tasks and devices, but the present thesis aims to provide a stable foundation for designing multilingual, inclusive VR reading experiences in human–computer interaction, language learning, and cross-cultural communication.

# Acknowledgement

No matter what we do in life, whether we can actually complete something does not depend on personal effort alone. I have always believed that it is the combined effect of 天時、地利、人和 – timing, environment, and people – that quietly shapes each outcome. The probability that all three align is, in fact, very small. Much like the complex interactions in my own research, truly "significant" combinations do not occur very often. Precisely because of this, I feel deeply grateful that I was able to complete this doctoral journey under such fortunate conditions.

**天時 (Heaven's timing).** People often say that after thirty-five it becomes harder to keep learning as freely as in one's twenties, so I feel very fortunate to have started my Ph.D. at Hokkaido University before turning thirty, and I hope I can go on learning new things while preserving my ability to think independently. I entered Hokkaido University as a research student in October 2022, just as generative AI tools such as ChatGPT began to appear and rapidly change research and daily life. By the time I formally enrolled in the doctoral program in April 2023, I had, in a sense, stepped into the "AI era" from its very beginning. These tools have brought great convenience, but they have also given me moments of some anxiety about whether humans might one day be "replaced." Through constant trial and adjustment, I have come to see that my own thinking must remain at the center, and that AI is best treated as a powerful tool that can support, but never replace, human insight. The training I received during the Ph.D. has strengthened this reflective way of thinking, and for that I am sincerely grateful to this particular moment in history.

**地利 (Favorable environment).** Japan has been, since my childhood, the country I most wanted to visit besides China. Whether it was 二次元 or broader aspects of Japanese culture, it is hard to summarize my reasons for wanting to come here in just a few sentences. I still remember, as a junior high school student who had not yet learned the Japanese syllabaries, pausing my MP3 player line by line and copying Japanese lyrics and anime lines into notebooks using Chinese phonetic approximations. In short, being able to study in Japan has truly been a stroke of luck in my life.

I would like to thank the Japanese government for unconditionally providing cost-of-living subsidies to all households, including foreign residents, even during this period of rising prices. I am also deeply grateful to the Ambitious and ExEx programs for their generous research and living support throughout these three years. Without such support, if I had needed to spend several hours every day doing part-time work just to maintain a basic livelihood, it would have been far more difficult to concentrate on my

research, and continuing in the Ph.D. program would have become considerably more challenging.

**人和 (Harmonious people).** Looking back on this journey, I truly feel that I have been very fortunate to meet kind and warm people along the way.

First and foremost, I would like to express my deepest gratitude to my supervisor, 坂本大介教授 (Prof. Daisuke Sakamoto). Over the past three years, Prof. Sakamoto has provided meticulous and patient guidance at every stage of my research: from defining the topic and designing experiments, to implementing systems, analyzing data, and writing and revising papers. When I first entered the lab, I was still unfamiliar with VR research; now I can independently design experiments and complete academic writing. This progress would not have been possible without his careful explanations, concrete suggestions, and gentle but steady encouragement. Before each entrance presentation, research proposal presentation, conference talk, and doctoral examination presentation, Prof. Sakamoto always took time from his busy schedule to go through my slides and scripts page by page, helping me to refine both the structure and the wording. Thanks to this, I was able to face each important presentation with a bit more confidence, even under great pressure. He has not only taught me how to conduct rigorous research, but also how to organize my thoughts, express my ideas clearly, and move forward step by step. Being able to meet such a patient, strict, and yet warm supervisor during my Ph.D. is one of the greatest fortunes in my life.

I am also sincerely grateful to all my labmates. The way everyone devotes themselves to research, the respectful and inclusive atmosphere, and the relatively free and relaxed daily environment have always supported me quietly in the background. When I first joined the lab, my Japanese was still very clumsy. 篠原さん and 菊地さん, who has already graduated, often chatted with me patiently and kindly. They gave me a lot of courage and very concretely helped my Japanese improve. 兼古さん is always very friendly and deeply interested in Chinese culture. I am often struck by how much he knows about culture, history, and information science, and by how quickly he comes up with good ways to deal with any problem. I have also learned a great deal from 崔さん, who has already graduated and had a very solid research foundation and often inquired about my research progress and offered thoughtful advice, from 阿部さん's rare combination of positivity and objectivity toward research, and from 日下部さん and 鈴木さん's solid research skills. Their examples have quietly encouraged and motivated me throughout these years.

I would also like to extend special thanks to all the volunteers who participated in the experiments in this dissertation. It is only because of their time and patient cooperation that the experimental parts of this thesis could be completed smoothly. In addition, I would like to thank the administrative staff at IST who helped me handle procedures again and again and who were always willing to offer advice and support. Their help ensured that I did not feel completely lost in a foreign administrative and cultural environment.

For daily life and emotional support, I owe my deepest thanks to my parents. They have continued to support my academic pursuits from afar while taking care of their own health and looking after my grandparents during my doctoral studies. Their encouragement has allowed me to concentrate on my research without constantly worrying about home. I am also deeply grateful to my husband, Dr. Zhiping Bian(卞志平), who came to Japan with me and has built our life here together as a family. His companionship, practical help, allowed me to lead a full and steady life during these doctoral years. Finally, I would like to thank Kang Kang(康康), Angel Cat born in October 2022, who has accompanied me throughout this journey and silently shared both the stressful and the peaceful days.

In closing, I would like to return once more to 天時、地利、人和. Personal effort is undoubtedly important, but without the complex and often accidental interplay of these three factors, I could not have come this far. For that reason, I will always try to remain humble and grateful, and never take any “smooth” outcome for granted. All I can do is to keep striving to become a better person, to maintain capacity for independent thought, to pass on to others the kindness and support that have warmed me, and, within the limits of my abilities, to contribute to future exchange and friendship between China and Japan.

# Bibliography


[1] David T Neal, Wendy Wood, Mengju Wu, and David Kurlander. The pull of the past: when do habits persist despite conflict with motives? *Personality and social psychology bulletin*, Vol. 37, No. 11, pp. 1428–1437, 2011.

[2] Jonathan St BT Evans and Keith E Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, Vol. 8, No. 3, pp. 223–241, 2013.

[3] Richard H Thaler, Cass R Sunstein, and John P Balz. Choice architecture. *The behavioral foundations of public policy*, Vol. 25, pp. 428–439, 2013.

[4] Naoto Fukasawa Design. About. Official website (in Japanese).

[5] Steve Krug. *Don't Make Me Think, Revisited: A Common Sense Approach to Web Usability*. New Riders, 3 edition, 2014.

[6] Jakob Nielsen. Enhancing the explanatory power of usability heuristics. In *Proceedings of CHI '94*, pp. 152–158, 1994.

[7] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson, 6th edition, 2021.

[8] Larry R. Squire and John T. Wixted. The cognitive neuroscience of human memory since h.m. *Annual Review of Neuroscience*, Vol. 34, pp. 259–288, 2011.

[9] Floris P De Lange, Micha Heilbron, and Peter Kok. How do expectations shape perception? *Trends in cognitive sciences*, Vol. 22, No. 9, pp. 764–779, 2018.

[10] Miles A Tinker and Donald G Paterson. Influence of type form on speed of reading. *Journal of Applied Psychology*, Vol. 12, No. 4, p. 359, 1928.

[11] Miles A Tinker and Donald G Paterson. Studies of typographical factors influencing speed of reading. iii. length of line. *Journal of Applied Psychology*, Vol. 13, No. 3, p. 205, 1929.

[12] MA Tinker and DG Paterson. Studies of typographical factors influencing speed of reading. xiii. methodological considerations. *Journal of Applied Psychology*, Vol. 20, No. 1, p. 132, 1936.

[13] M.A. Tinker. *Legibility of Print*. Iowa State University Press, 1963.

[14] Andrew Dillon. Reading from paper versus screens: A critical review of the empirical literature. *Ergonomics*, Vol. 35, No. 10, pp. 1297–1326, 1992.

[15] Apple Computer, Inc. *Inside Macintosh: Imaging With QuickDraw*, 1994.

[16] Iso 9241-12: Ergonomic requirements for office work with visual display terminals (vdts) — part 12: Presentation of information. Technical report, International Organization for Standardization, 1998.

[17] J.D. Berry and J. Hudson. *Now Read this: The Microsoft Cleartype Font Collection*. Microsoft, 2004.

[18] Apple Inc. iphone 4 — technical specifications, 2010.

[19] Aries Arditi and Jianna Cho. Serifs and font legibility. *Vision Research*, Vol. 45, No. 23, pp. 2926–2933, 2005.

[20] Microsoft. Verdana font family — microsoft typography, 2025.

[21] Phillip Guan, Eric Penner, Joel Hegland, Benjamin Letham, and Douglas Lanman. Perceptual requirements for world-locked rendering in ar and vr. In *SIGGRAPH Asia 2023 Conference Papers*, pp. 1–10, 2023.

[22] Chumin Zhao, Andrea S Kim, Ryan Beams, and Aldo Badano. Spatiotemporal image quality of virtual reality head mounted displays. *Scientific Reports*, Vol. 12, No. 1, p. 20235, 2022.

[23] Marissa Howard Lynn, Gang Luo, Matteo Tomasi, Shrinivas Pundlik, and Kevin E. Houston. Measuring virtual reality headset resolution and field of view: Implications for vision care applications. *Optometry and Vision Science*, Vol. 97, No. 8, pp. 573–582, 2020.

[24] Jianghao Xiong, En-Lin Hsiang, Ziqian He, Tao Zhan, and Shin-Tson Wu. Augmented reality and virtual reality displays: emerging technologies and future perspectives. *Light: Science & Applications*, Vol. 10, No. 1, p. 216, 2021.

[25] Oculus VR, Inc. *Oculus VR Best Practices Guide*. Oculus VR, 2014. Design and rendering guidance for comfortable VR (distortion correction, FoV/PPD considerations, text clarity).

[26] Daniel Pohl, Gregory S. Johnson, and Timo Bolkart. Improved pre-warping for wide angle, head mounted displays. VRST '13, p. 259–262, New York, NY, USA, 2013. Association for Computing Machinery.

[27] Markku Kilpeläinen and Jukka Häkkinen. An effective method for measuring text legibility in xr devices reveals clear differences between three devices. *Frontiers in Virtual Reality*, Vol. 4, p. 1243387, 2023.

[28] Rufat Rzayev, Polina Ugnivenko, Sarah Graf, Valentin Schwind, and Niels Henze. Reading in vr: The effect of text presentation type and location. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–10, 2021.

[29] Maxime Cauz, Antoine Clarinval, and Bruno Dumas. Text readability in augmented reality: a multivocal literature review. *Virtual Reality*, Vol. 28, No. 1, p. 59, 2024.

[30] Charles Perfetti. Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, Vol. 11, No. 4, pp. 357–383, 2007.

[31] Keith E Stanovich. Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading research quarterly*, pp. 32–71, 1980.

[32] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, Vol. 124, No. 3, p. 372, 1998.

[33] Uschi Cop, Denis Drieghe, and Wouter Duyck. Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PloS one*, Vol. 10, No. 8, p. e0134008, 2015.

[34] Tien-Yu Ho, Hao-Chuan Wang, and Shong-Hong Lai. Non-native language reading support with display of machine translation based on eye-tracking and sentence-level mapping. In *Proceedings of the Sixth International Symposium of Chinese CHI*, ChineseCHI '18, p. 57–63, New York, NY, USA, 2018. Association for Computing Machinery.

[35] Shaun Wallace, Zoya Bylinskii, Jonathan Dobres, Bernard Kerr, Sam Berlow, Rick Treitman, Nirmal Kumawat, Kathleen Arpin, Dave B. Miller, Jeff Huang, and Ben D. Sawyer. Towards individuated reading experiences: Different fonts increase reading speed for different individuals. *ACM Trans. Comput.-Hum. Interact.*, Vol. 29, No. 4, March 2022.

[36] Ram Frost. Towards a universal model of reading. *Behavioral and brain sciences*, Vol. 35, No. 5, pp. 263–279, 2012.

[37] Charles A Perfetti and Li Hai Tan. The time course of graphic, phonological, and semantic activation in chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 24, No. 1, p. 101, 1998.

[38] Su-Ling Yeh and Jing-Ling Li. Role of structure and component in judgments of visual similarity of chinese characters. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 28, No. 4, p. 933, 2002.

[39] Jonathan Grainger. Letters, words, sentences, and reading. *Journal of Cognition*, Vol. 7, No. 1, p. 66, 2024.

[40] I-Fan Su, Sin-Ching Cassie Mak, Lai-Ying Milly Cheung, and Sam-Po Law. Taking a radical position: evidence for position-specific radical representations in chinese character recognition using masked priming erp. *Frontiers in Psychology*, Vol. 3, p. 333, 2012.

[41] Hui Wang and Gordon E Legge. Comparing the minimum spatial-frequency content for recognizing chinese and alphabet characters. *Journal of vision*, Vol. 18, No. 1, pp. 1–1, 2018.

[42] Yasushi Hino, Shinobu Miyamura, and Stephen J Lupker. The nature of orthographic–phonological and orthographic–semantic relationships for japanese kana and kanji words. *Behavior research methods*, Vol. 43, No. 4, pp. 1110–1151, 2011.

[43] Mutsuo Ijuin and Taeko N Wydell. A reading model from the perspective of japanese orthography: Connectionist approach to the hypothesis of granularity and transparency. *Journal of Learning Disabilities*, Vol. 51, No. 5, pp. 490–498, 2018.

[44] Taeko Nakayama Wydell and Brian Butterworth. A case study of an english-japanese bilingual with monolingual dyslexia. *Cognition*, Vol. 70, No. 3, pp. 273–305, 1999.

[45] Melanie Labusch, Pablo Gómez, and Manuel Perea. Does adding an accent mark hinder lexical access? evidence from spanish. *Journal of Cultural Cognitive Science*, Vol. 6, No. 2, pp. 219–228, 2022.

[46] Denisa Bordag, Andreas Opitz, Max Polter, and Michael Meng. Non-native readers are more sensitive to changes in surface linguistic information than native readers. *Bilingualism: Language and cognition*, Vol. 24, No. 4, pp. 599–611, 2021.

[47] Aurélie Calabrese, Allen MY Cheong, Sing-Hang Cheung, Yingchen He, MiYoung Kwon, J Stephen Mansfield, Ahalya Subramanian, Deyue Yu, and Gordon E Legge. Baseline mnread measures for normally sighted subjects from childhood to old age. *Investigative ophthalmology & visual science*, Vol. 57, No. 8, pp. 3836–3843, 2016.

[48] Oxford Research Encyclopedia of Education. Bilingual Education. `https://oxfordre.com/education/display/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-1059`, 2021. Defines L1 as a person's native language and L2 as the target/second language. Accessed 2025-09-07.

[49] Matti Rahkonen and Taina Juurakko. Logit analysis in l2 research: measuring l1 and l2/ln effects. *International Journal of Applied Linguistics*, Vol. 8, No. 1, pp. 81–110, 1998.

[50] Kun Yin, En-Lin Hsiang, Junyu Zou, Yannanqi Li, Zhiyong Yang, Qian Yang, Po-Cheng Lai, Chih-Lung Lin, and Shin-Tson Wu. Advanced liquid crystal devices for augmented reality and virtual reality displays: principles and applications. *Light: Science & Applications*, Vol. 11, No. 1, p. 161, 2022.

[51] Mike Alger. Visual design methods for virtual reality. *Ravensbourne. http://aperturesciencellc. com/vr/VisualDesignMethodsforVR_MikeAlger. pdf*, 2015.

[52] Wikipedia contributors. Kana. `https://en.wikipedia.org/wiki/Kana`, 2025. Accessed: 2025-12-11.

[53] Takumi Kosaka. The effects of chunk reading training on the syntactic processing skills and reading spans of japanese learners of english. 2023.

[54] Stanislas Dehaene, Felipe Pegado, Lucia W Braga, Paulo Ventura, Gilberto Nunes Filho, Antoinette Jobert, Ghislaine Dehaene-Lambertz, Régine Kolinsky, José Morais, and Laurent Cohen. How learning to read changes the cortical networks for vision and language. *science*, Vol. 330, No. 6009, pp. 1359–1364, 2010.

[55] Ziteng Wang, Pengfei Gao, Liang Ma, and Wei Zhang. Effects of font size, line spacing, and font style on legibility of chinese characters on consumer-based virtual reality displays. In *HCI International 2020–Late Breaking Posters: 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22*, pp. 468–474. Springer, 2020.

[56] Xiaozhou Zhou, Yixue Wang, Ziwei Zhang, Xu-Yi Qiu, and Yuqing Zhou. Research on the legibility of chinese display character sizes in virtual environments. *Displays*, Vol. 81, p. 102589, 2024.

[57] Jean-Baptiste Bernard and Susana TL Chung. The dependence of crowding on flanker complexity and target–flanker similarity. *Journal of vision*, Vol. 11, No. 8, pp. 1–1, 2011.

[58] Najib J Majaj, Denis G Pelli, Peri Kurshan, and Melanie Palomares. The role of spatial frequency channels in letter identification. *Vision research*, Vol. 42, No. 9, pp. 1165–1184, 2002.

[59] Vladimir Y Vildavski, Luca Lo Verde, Gail Blumberg, Joss Parsey, and Anthony M Norcia. Pseudosloan: A perimetric-complexity and area-controlled font for vision and reading research. *Journal of Vision*, Vol. 22, No. 10, pp. 7–7, 2022.

[60] Agency for Cultural Affairs. List of commonly used kanji (jōyō kanji). `https://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/kijun/naikaku/kanji/`. Cabinet notification, November 30, 2010. (In Japanese). Accessed: 2025-12-11.

[61] General Standard Chinese Character Dataset. General standard chinese characters with pinyin. https://github.com/ben-hua/general_standard_chinese/blob/main/gsc_pinyin.csv. GitHub repository. Accessed: 2025-12-11.

[62] Pillow developers. Pillow: Python imaging library. https://pypi.org/project/pillow/. Python image-processing library. Accessed: 2025-12-11.

[63] The Editors of Encyclopædia Britannica. Kana. https://www.britannica.com/topic/kana, 2025. Encyclopædia Britannica. Accessed 2025-09-07.

[64] Jukka Hyönä, Lei Cui, Timo T. Heikkilä, Birgitta Paranko, Yun Gao, and Xingzhi Su. Reading compound words in finnish and chinese: An eye-tracking study. *Journal of Memory and Language*, Vol. 134, p. 104474, 2024.

[65] Gordon E. Legge. *Psychophysics of Reading in Normal and Low Vision.* CRC Press, 2006.

[66] Alexei Tarasov, Marius Pedersen, and Peter Østvand. Factors influencing the perception of font legibility in the context of geographic maps. *Cartography and Geographic Information Science*, Vol. 42, No. Suppl 1, pp. Sup1:50–60, 2015.

[67] Yuki Matsuura, Tsutomu Terada, Tomohiro Aoki, Susumu Sonoda, Naoya Isoyama, and Masahiko Tsukamoto. Readability and legibility of fonts considering shakiness of head mounted displays. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers*, ISWC '19, pp. 150–157, New York, NY, USA, 2019. ACM.

[68] Tilman Dingler, Kai Kunze, and Benjamin Outram. VR Reading UIs: Assessing Text Parameters for Reading in VR. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pp. LBW094:1–6, New York, NY, USA, 2018. ACM.

[69] Chloe Grout, William Rogers, Mark Apperley, and Steve Jones. Reading text in an immersive head-mounted display: An investigation into displaying desktop interfaces in a 3d virtual environment. In *Proceedings of the 15th New Zealand Conference on Human-Computer Interaction*, CHINZ '15, pp. 9–16, New York, NY, USA, 2015. ACM.

[70] Zhengyin Gu, Xinle Bao, Ying Zhu, Qijun Wang, Wei Gao, Duming Wang, Yu Tian, and You Li. Rejecting small! accepting larger? optimizing the font size of chinese characters basing legibility and visual fatigue in ost-hmds. *Ergonomics*, Vol. 67, No. 12, pp. 2030–2043, 2024.

[71] Hui Wang, Xuanzi He, and Gordon E Legge. Effect of pattern complexity on the visual span for chinese and alphabet characters. *Journal of vision*, Vol. 14, No. 8, pp. 6–6, 2014.

[72] Google. Noto sans english. https://fonts.google.com/noto/specimen/Noto+Sans, 2007. Google Fonts. Accessed: February 26, 2007.

[73] Google. Noto sans japanese. https://fonts.google.com/noto/specimen/Noto+Sans+JP?query=noto+sans, 2007. Google Fonts. Accessed: February 26, 2007.

[74] Google. Noto sans chinese. https://fonts.google.com/noto/specimen/Noto+Sans+SC?query=noto+sans+, 2007. Google Fonts. Accessed: February 26, 2007.

[75] Google. Noto serif english. https://fonts.google.com/noto/specimen/Noto+Serif?query=noto+serif, 2007. Google Fonts. Accessed: February 26, 2007.

[76] Google. Noto serif japanese. https://fonts.google.com/noto/specimen/Noto+Serif+JP?query=noto+serif, 2007. Google Fonts. Accessed: February 26, 2007.

[77] Google. Noto serif chinese. https://fonts.google.com/noto/specimen/Noto+Serif+SC?query=noto+serif, 2007. Google Fonts. Accessed: February 26, 2007.

[78] Donald T Campbell and Donald W Fiske. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin*, Vol. 56, No. 2, p. 81, 1959.

[79] Herbert W. Marsh. Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, Vol. 13, No. 4, pp. 335–361, 1989.

[80] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[81] Bill Albert and Tom Tullis. *Measuring the user experience: Collecting, analyzing, and presenting UX metrics*. Morgan Kaufmann, 2022.

[82] Sarah H Creem-Regehr, Jeanine K Stefanucci, and Bobby Bodenheimer. Perceiving distance in virtual reality: theoretical insights from contemporary technologies. *Philosophical Transactions of the Royal Society B*, Vol. 378, No. 1869, p. 20210456, 2023.

[83] Sabina Baceviciute, Thomas Terkildsen, and Georgios Makransky. Remediating learning from non-immersive to immersive media: Using eeg to investigate the effects of environmental embeddedness on reading in virtual reality. *Computers & Education*, Vol. 164, p. 104122, 2021.

[84] Matti Pölönen, Jussi Takatalo, Göte Nyman, Miikka Laitinen, and Gábor Bárdy. Individual differences in subjective experiences and eye strain during 3d tv viewing. *Electronic Imaging*, Vol. 2013, No. 1, p. Journal of Electronic Imaging 011005, 2013.

[85] Jun Zhang, Jingbo Liu, Srichand Jasti, Rajaraman Suryakumar, and Mark A. Bullimore. Visual demand and acuity reserve of chinese versus english newspapers. *Optometry and Vision Science*, Vol. 97, No. 10, pp. 865–870, 2020.

[86] Adam Glazier and Chris McKenzie. Designing screen interfaces for vr (google i/o '17). `https://www.youtube.com/watch?v=ES9jArHRFHQ`, 2017. Google I/O 2017 session video.

[87] Google. Daydream app quality requirements – design. `https://developers.google.com/vr/distribute/daydream/design-requirements`, 2024. UX-D1: Minimum angular size guidance for legible text in Daydream.

[88] Tanja Kojić, Danish Ali, Robert Greinacher, Sebastian Möller, and Jan-Niklas Voigt-Antons. User experience of reading in virtual reality—finding values for text distance, size and contrast. In *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6. IEEE, 2020.

[89] Takashi Shibata, Joohwan Kim, David M. Hoffman, and Martin S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, Vol. 11, No. 8, p. 11, 2011.

[90] Sina Masnadi, Kevin P. Pfeil, Jose-Valentin T. Sera-Josef, and Joseph J. LaViola Jr. Field of view effect on distance perception in virtual reality. In *Proceedings of the 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 542–543. IEEE, 2021.

[91] Michael L. Bernard, Chia Hui Liao, and Melissa M. Mills. The effects of font type and size on the legibility and reading time of online text by older adults. In *CHI'01 Extended Abstracts on Human Factors in Computing Systems*, pp. 175–176. ACM, 2001.

[92] Katsumi Minakata and Sofie Beier. The dispute about sans serif versus serif fonts: An interaction between the variables of serif and stroke contrast. *Acta Psychologica*, Vol. 228, p. 103623, 2022.

[93] Sofie Beier and Chiron A. T. Oderkerk. Smaller visual angles show greater benefit of letter boldness than larger visual angles. *Acta Psychologica*, Vol. 199, p. 102904, 2019.

[94] Sara Vecino, Jonas Mehtali, Javier de Andrés, Martin Gonzalez-Rodriguez, and Daniel Fernandez-Lanvin. How does serif vs sans serif typeface impact the usability of e-commerce websites? *PeerJ Computer Science*, Vol. 8, p. e1139, 2022.

[95] Lili Yu, Qiaoming Zhang, Caspian Priest, Erik D Reichle, and Heather Sheridan. Character-complexity effects in chinese reading and visual search: A comparison and theoretical implications. *Quarterly Journal of Experimental Psychology*, Vol. 71, No. 1, pp. 140–151, 2018.

[96] Miia Sainio, Jukka Hyönä, Kazuo Bingushi, and Raymond Bertram. The role of interword spacing in reading japanese: An eye movement study. *Vision research*, Vol. 47, No. 20, pp. 2575–2584, 2007.

[97] Yuqi Hao and Ming Yan. Is direction-specific reading experience transferrable across orthographies? eye tracking evidence from chinese-japanese bilinguals. *Scientific Studies of Reading*, Vol. 29, No. 3, pp. 290–302, 2025.

[98] Huixian Qiao, Ying Cao, Zhihan Sun, Ziwei Yang, Dingyi Li, and Shihua Li. Eye-tracking in recent l2 learner research. In *2022 3rd International Conference on Mental Health, Education and Human Development (MHEHD 2022)*, pp. 1274–1281. Atlantis Press, 2022.

[99] Natsumi Kajii, Tatjana A. Nazir, and Naoyuki Osaka. Eye movement control in reading unspaced text: the case of the japanese script. *Vision Research*, Vol. 41, No. 19, pp. 2503–2510, 2001.

[100] Chi Yui Leung, Masatoshi Sugiura, Daisuke Abe, Lisa Yoshikawa, et al. The perceptual span in second language reading: An eye-tracking study using a gaze-contingent moving window paradigm. *Open Journal of Modern Linguistics*, Vol. 4, No. 05, p. 585, 2014.

[101] Yaqiong Cui. Eye movements of second language learners when reading spaced and unspaced chinese texts. *Frontiers in Psychology*, Vol. 14, p. 783960, 2023.

[102] Yevgeni Berzak, Chie Nakamura, Amelia Smith, Emily Weng, Boris Katz, Suzanne Flynn, and Roger Levy. Celer: A 365-participant corpus of eye movements in l1 and l2 english reading. *Open Mind*, Vol. 6, pp. 41–50, 2022.

[103] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, Vol. 8, No. 3, p. 33, 2008.

[104] Inc. Meta Platforms. Dynamic resolution, 2025.

[105] Yorick Wilks and Dann Fass. The preference semantics family. *Computers & Mathematics with Applications*, Vol. 23, No. 2-5, pp. 205–221, 1992.

[106] Xiaozhou Zhou, Yixue Wang, Ziwei Zhang, Xu-Yi Qiu, and Yuqing Zhou. Research on the legibility of chinese display character sizes in virtual environments. *Displays*, Vol. 81, p. 102589, 2024.

[107] Python Software Foundation. Python 3 Documentation. `https://docs.python.org/3/`, 2025. Accessed 2025-09-07.

[108] pandas development team. pandas: Documentation. `https://pandas.pydata.org/docs/`, 2025. Accessed 2025-09-07.

[109] NumPy Developers. NumPy Documentation. `https://numpy.org/doc/`, 2025. Accessed 2025-09-07.

[110] Keith Rayner, Timothy J Slattery, and Nathalie N Bélanger. Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review*, Vol. 17, No. 6, pp. 834–839, 2010.

[111] Sixin Liao, Erik D. Reichle, et al. Dynamic reading in a digital age: New insights on cognition. *Trends in Cognitive Sciences*, Vol. 28, No. 1, pp. 43–55, 2024.

[112] Naoyuki Osaka and Koichi Oda. Effective visual field size necessary for vertical reading during japanese text processing. *Bulletin of the Psychonomic Society*, Vol. 29, No. 4, pp. 345–347, 1991.

[113] Albrecht Werner Inhoff and Weimin Liu. The perceptual span and oculomotor activity during the reading of chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 24, No. 1, p. 20, 1998.

[114] Maxim Likhanov, Olga Bogdanova, Evgenia Alenina, Tatiana Kolienko, and Yulia Kovas. No evidence of a positive effect of learning chinese language as an l2 on spatial ability. *Scientific Reports*, Vol. 13, No. 1, p. 1262, 2023.

[115] Ronald A Rensink. The dynamic representation of scenes. *Visual cognition*, Vol. 7, No. 1-3, pp. 17–42, 2000.

[116] Lin Chen, Charles Perfetti, Ying Leng, and You Li. Word superiority effect for native chinese readers and low-proficiency chinese learners. *Applied Psycholinguistics*, Vol. 39, No. 6, pp. 1097–1115, 2018.

[117] Johannes Lang and Miguel A Nacenta. Perception of letter glyph parameters for infotypography. *ACM Transactions on Graphics (TOG)*, Vol. 41, No. 4, pp. 1–21, 2022.

[118] Li Hai Tan, Angela R Laird, Karl Li, and Peter T Fox. Neuroanatomical correlates of phonological processing of chinese characters and alphabetic words: A meta-analysis. *Human brain mapping*, Vol. 25, No. 1, pp. 83–91, 2005.

[119] Singhal and Meena. A comparison of l1 and l2 reading: Cultural differences and schema. *The internet TESL journal*, Vol. 4, No. 10, pp. 4–10, 1998.

[120] Evelien Akker and Anne Cutler. Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, Vol. 6, No. 2, pp. 81–96, 2003.

[121] Baker and Rachel E. Non-native perception of native english prominence. In *Proceedings of the 5th International Conference on Speech Prosody*, Vol. 100171, pp. 1–4. Cambridge University Press, 2010.

[122] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R Doyle, Leigh Clark, and Benjamin R Cowan. See what i' m saying? comparing intelligent personal assistant use for native and non-native language speakers. In *22nd international conference on human-computer interaction with mobile devices and services*, pp. 1–9, 2020.

[123] Saya Kawase, Chris Davis, and Jeesun Kim. Impact of japanese l1 rhythm on english l2 speech. *Language and Speech*, p. 00238309241247210, 2024.

[124] Eun-Kyung Lee and Scott Fraundorf. Native-like processing of prominence cues in l2 written discourse comprehension: Evidence from font emphasis. *Applied Psycholinguistics*, Vol. 40, No. 2, pp. 373–398, 2019.

[125] Hanna S Gauvin and Jan H Hulstijn. Exploring a new technique for comparing bilinguals' l1 and l2 reading speed. *Reading in a Foreign Language*, 2010.

[126] Benedetta Bassetti. Effects of adding interword spacing on chinese reading: A comparison of chinese native readers and english readers of chinese as a second language. *Applied Psycholinguistics*, Vol. 30, No. 4, pp. 757–775, 2009.

[127] Leigh B Fernandez, Ricarda Bothe, and Shanley EM Allen. The role of l1 reading direction on l2 perceptual span: An eye-tracking study investigating hindi and urdu speakers. *Second Language Research*, Vol. 39, No. 2, pp. 447–469, 2023.

[128] Y Fu, SP Liversedge, X Bai, M Moosa, and C Zang. Character representations i n lexical identification during natural reading. *Manuscript submitted for publication*, 2023.

[129] Simon P Liversedge, Henri Olkoniemi, Chuanli Zang, Xin Li, Guoli Yan, Xuejun Bai, and Jukka Hyönä. Universality in eye movements and reading: A replication with increased power. *Cognition*, Vol. 242, p. 105636, 2024.

[130] Hye K Pae and Yong-Won Lee. The resolution of visual noise in word recognition. *Journal of psycholinguistic research*, Vol. 44, pp. 337–358, 2015.

[131] Chiron AT Oderkerk and Sofie Beier. Fonts of wider letter shapes improve letter recognition in parafovea and periphery. *Ergonomics*, Vol. 65, No. 5, pp. 753–761, 2022.

[132] Katsumi Minakata and Sofie Beier. The effect of font width on eye movements during reading. *Applied ergonomics*, Vol. 97, p. 103523, 2021.

[133] Ulrike Jessner. A dst model of multilingualism and the role of metalinguistic awareness. *The modern language journal*, Vol. 92, No. 2, pp. 270–283, 2008.

[134] Yuki Matsuura, Tsutomu Terada, Tomohiro Aoki, Susumu Sonoda, Naoya Isoyama, and Masahiko Tsukamoto. Readability and legibility of fonts considering

shakiness of head mounted displays. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 150–159, 2019.

[135] Jacek Jankowski, Krystian Samp, Izabela Irzynska, Marek Jozwowicz, and Stefan Decker. Integrating text with video and 3d graphics: The effects of text drawing styles on text readability. In *proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1321–1330, 2010.

[136] Dimitrios Saredakis, Ancret Szpak, Brandon Birckhead, Hannah A. D. Keage, Albert Rizzo, and Tobias Loetscher. Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. *Frontiers in Human Neuroscience*, Vol. Volume 14 - 2020, , 2020.

[137] Huidan Zhang. Supporting data and materials for vr glyph legibility experiments. `https://doi.org/10.17605/OSF.IO/FJSTU`. Open Science Framework (OSF) repository. Accessed: 2025-12-11.

[138] Jingjing Li. Reception of audiovisual translation in china: a survey study. *Perspectives*, pp. 1–21, 2024.

[139] Shuting Huang, Yacong Zhou, Menglin Du, Ruiming Wang, and Zhenguang G. Cai. Character amnesia in chinese handwriting: a mega-study analysis. *Language Sciences*, Vol. 85, p. 101383, 2021.

[140] Yueyue Liu, Deyang Zhang, Yongli Fu, Yanlou Liu, and Wenguang He. Phonological and orthographic processing during second language typing production of chinese-english bilinguals. *Acta Psychologica*, Vol. 255, p. 104910, 2025.

[141] Baoning Zhong and Yeqin Kang. Chinese efl teachers' perception and practice of phonics instruction. *Journal of Language Teaching and Research*, Vol. 12, No. 6, pp. 990–999, 2021.

[142] Rachel Kraut. The relationship between morphological awareness and morphological decomposition among english language learners. *Reading and Writing*, Vol. 28, pp. 873–890, 2015.

[143] Wayne A. Wickelgren. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, Vol. 41, No. 1, pp. 67–85, 1977.

[144] Fangyi Li, Qiang Shen, Ying Li, and Neil Mac Parthaláin. Handwritten chinese character recognition using fuzzy image alignment. *Soft Computing*, Vol. 20, pp. 2939–2949, 2016.

[145] Pinterest board. Examples of chinese characters with embedded latin letters and graphical simplifications. `https://jp.pinterest.com/pin/392024342562295638/`. Design examples of stylized Hanzi/Kanji. Accessed: 2025-12-11.

[146] Felix Leminen and Shurui Bai. Applying visual mnemonics enhances chinese characters learning for chinese as second language learners: A mixed-method study. In *Proceedings of the 7th International Conference on Digital Technology in Education*, ICDTE '23, p. 150–154, New York, NY, USA, 2023. Association for Computing Machinery.

[147] Chung-mou Cheng. Perception of chinese characters. *Acta Psychologica Taiwanica*, Vol. 23, No. 2, pp. 137–153, 1981.

[148] Xiuqi Chen, Yucheng Lu, and Jie Liu. Stroke-speech: A multi-channel input method for chinese characters. In *Proceedings of the Eighth International Workshop of Chinese CHI*, Chinese CHI '20, p. 69–72, New York, NY, USA, 2020. Association for Computing Machinery.

[149] Sofie Beier and Chiron AT Oderkerk. High letter stroke contrast impairs letter recognition of bold fonts. *Applied Ergonomics*, Vol. 97, p. 103499, 2021.

[150] Sebastian P Korinth, Kerstin Gerstenberger, and Christian J Fiebach. Wider letter-spacing facilitates word processing but impairs reading rates of fast readers. *Frontiers in psychology*, Vol. 11, p. 497641, 2020.

[151] Madoka Ohnishi and Koichi Oda. The effect of character stroke width on legibility: The relationship between duty ratio and contrast threshold. *Vision Research*, Vol. 185, pp. 1–8, 2021.

[152] Gary S Rubin, Mary Feely, Sylvie Perera, Katherin Ekstrom, and Elizabeth Williamson. The effect of font and line width on reading speed in people with mild to moderate vision loss. *Ophthalmic and Physiological Optics*, Vol. 26, No. 6, pp. 545–554, 2006.

[153] Luz Rello and Ricardo Baeza-Yates. The effect of font type on screen readability by people with dyslexia. *ACM Transactions on Accessible Computing (TACCESS)*, Vol. 8, No. 4, pp. 1–33, 2016.

[154] Maneerut Chatrangsan and Helen Petrie. The effect of typeface and font size on reading text on a tablet computer for older and younger people. In *Proceedings of the 16th International Web for All Conference*, pp. 1–10, 2019.

[155] Ahmad Zamzuri Mohamad Ali, Rahani Wahid, Khairulanuar Samsudin, and Muhammad Zaffwan Idris. Reading on the computer screen: Does font type have effects on web text readability?. *International Education Studies*, Vol. 6, No. 3, pp. 26–35, 2013.

[156] Aries Arditi and Jianna Cho. Serifs and font legibility. *Vision research*, Vol. 45, No. 23, pp. 2926–2933, 2005.

[157] Anna Rita Galiano, Vanessa Augereau-Depoix, Nicolas Baltenneck, Laura Latour, and Hind Drissi. Luciole, a new font for people with low vision. *Acta Psychologica*, Vol. 236, p. 103926, 2023.

[158] Neil Newbold and Lee Gillam. The linguistics of readability: The next step for word processing. In *Proceedings of the NAACL HLT 2010 workshop on computational linguistics and writing: Writing processes and authoring aids*, pp. 65–72, 2010.

[159] Sofie Beier. *Typeface legibility: towards defining familiarity*. Royal College of Art (United Kingdom), 2009.

[160] Nobuhiko Akamatsu. The effects of first language orthographic features on word recognition processing in english as a second language. *Reading and Writing*, Vol. 11, pp. 381–403, 1999.

[161] Huidan Zhang. Font experimental records and interview transcripts. `https://github.com/zhanghuidan666/Font-Experimental-record-.git`. GitHub repository containing anonymized interview records and experimental logs. Accessed: 2025-12-11.

[162] Stefan Blohm, Valentin Wagner, Matthias Schlesewsky, and Winfried Menninghaus. Sentence judgments and the grammar of poetry: Linking linguistic structure and poetic effect. *Poetics*, Vol. 69, pp. 41–56, 2018.

[163] Xin Gao, Jeroen Dera, Annabel D Nijhof, and Roel M Willems. Is less readable liked better? the case of font readability in poetry appreciation. *Plos One*, Vol. 14, No. 12, p. e0225757, 2019.

[164] Jo Mackiewicz. Audience perceptions of fonts in projected powerpoint text slides. *Technical communication*, Vol. 54, No. 3, pp. 295–307, 2007.

# Appendix A List of Research Achievements

## A.1 Journal Articles (peer-reviewed)

[1] Zhang, Huidan, Daisuke Sakamoto, and Tetsuo Ono. "Study on font preferences of native and non-native speakers in a virtual reality environment." Frontiers in Virtual Reality 6 (2025): 1590871.

[2] Zhang, Huidan, Daisuke Sakamoto, and Tetsuo Ono. "Understanding glyph characteristics on the legibility for native and non-native speakers in virtual reality." Frontiers in Virtual Reality 6 (2025): 1579525.

## A.2 International Conference Papers (peer-reviewed)

[3] Zhang, Huidan, Daisuke Sakamoto, and Tetsuo Ono. "Perceiving Multilingual Text in Virtual Reality: Glyph Complexity and Font Effects on Preferred Viewing Distance." Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology. 2025.

# Appendix B Power Analysis of GLM Models

This appendix reports post-hoc power analyses for the full-factorial GLM models based on the sample of 30 participants (10 native readers per language group). Because the experiments—especially the glyph-identification study in Chapter 4—include several visual factors and their interactions, there is a natural concern that the large number of terms may reduce statistical power and lead to imprecise estimation. To check the robustness of the effects discussed in the main text, we therefore conducted a set of Monte Carlo–based power analyses.

The full-factorial GLM specification is identical to that reported in Chapter 4, including *Native*, *Font*, and *Glyph* and all of their interactions. Among the three studies in this dissertation, the Chapter 4 glyph-identification experiment represents the most demanding case: the glyph factor is operationalized through two dimensions (complexity and similarity), while the other visual parameters match those used in the distance-adjustment and font-preference experiments. In this sense, Chapter 4 provides a conservative test of whether the sample size is sufficient for the designs used in this work.

For each of the three language experiments (Chinese, Japanese, English) and each viewing distance (0.5, 2.5, 5, and 10 m), a Monte Carlo simulation–based power analysis with 500 iterations was carried out. In each iteration, data were generated under the fitted GLM and the model was re-estimated; power for a given term was defined as the proportion of iterations in which that term reached statistical significance at $\alpha = .05$. The resulting estimated power values are visualized in Figure B.1,Figure B.2and FigureB.3. In all panels, the red dashed vertical line marks the conventional target power of 0.80.

Bars annotated with white labels correspond to effects that are both statistically significant in the GLM and emphasized in the main chapters. Bars annotated with purple labels mark effects that reached significance in the GLM but did not achieve the 0.80 power threshold and, in several cases, also showed very small effect sizes (for example, $\eta^2 < .01$ for some *Error* terms in the Japanese experiment). Such effects are treated as exploratory and are not interpreted in detail in the dissertation.

Overall, the figures indicate that the design has reasonable power for the primary main effects and a subset of theoretically relevant two-way interactions, especially in the Chinese and Japanese experiments, where visual factors (e.g., glyph complexity) produce comparatively large effects. In practice, the terms whose estimated power exceeds 0.80 largely overlap with the effects highlighted and interpreted in the main chapters. By

contrast, many higher-order interactions show more modest power and trivial effect sizes, and are therefore treated as exploratory. These checks support the decision in the main text to focus interpretation on robust main and two-way patterns and to approach complex, higher-order interactions with caution. The scripts and raw simulation outputs are available in the OSF repository [137] for transparency.

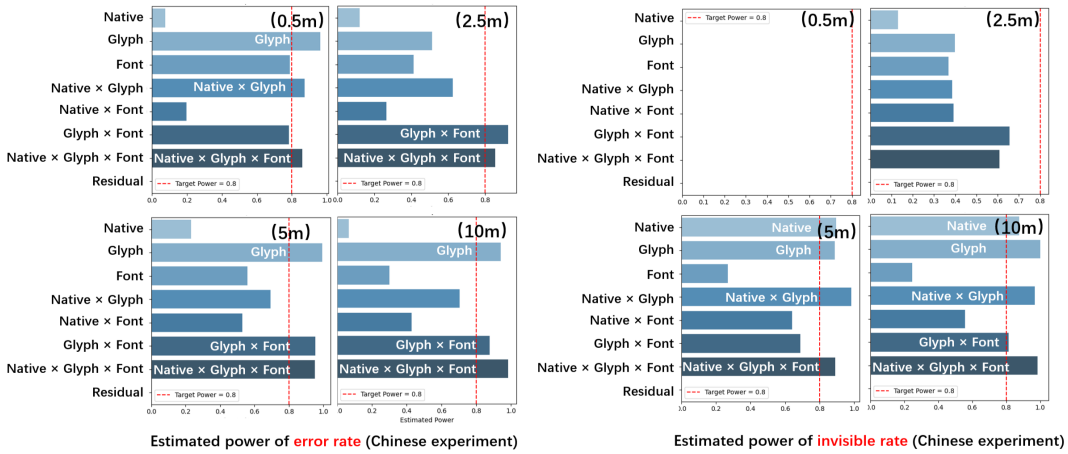



Fig. B.1: Simulation-based power analysis for the Chinese experiment. Bars show estimated power for each GLM term in the error (left) and invisibility (right) models; the red dashed line marks the target power of 0.80. White labels mark effects discussed in the main text.

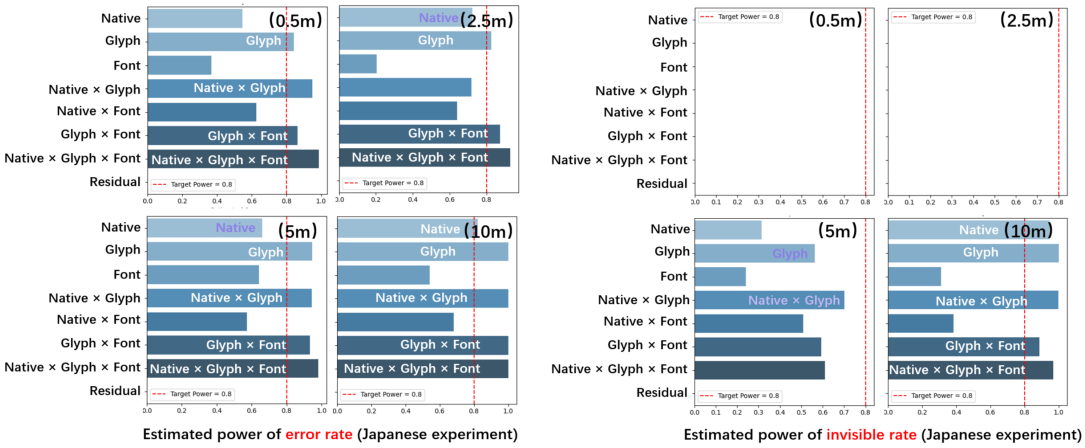



Fig. B.2: Simulation-based power analysis for the Japanese experiment. Bars show estimated power for each GLM term in the error (left) and invisibility (right) models; the red dashed line marks the target power of 0.80. White labels mark effects discussed in the main text, purple labels mark exploratory effects.

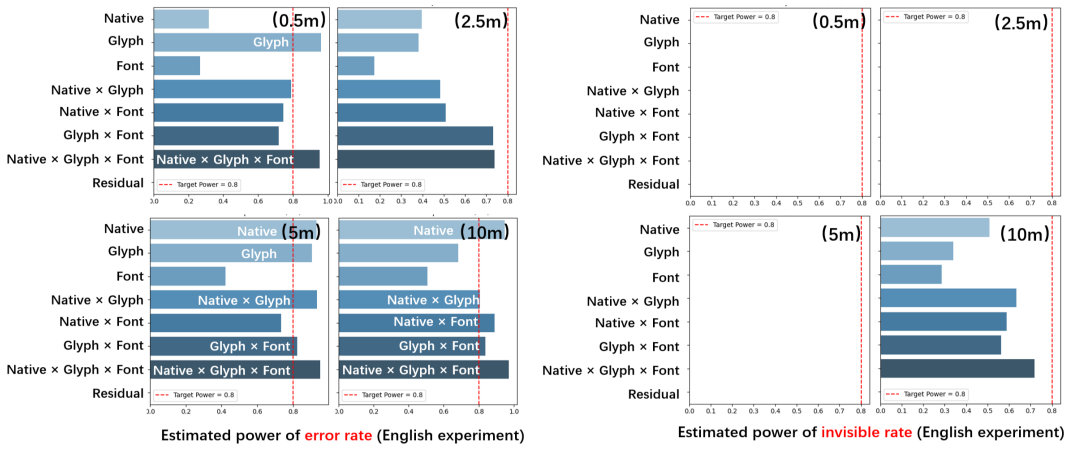



Fig. B.3: Simulation-based power analysis for the English experiment. Bars show estimated power for each GLM term in the error (left) and invisibility (right) models; the red dashed line marks the target power of 0.80. White labels mark effects discussed in the main text.

# List of Figures


1.1 Structure of the dissertation . . . . . 8

2.1 Overview of the VR system . . . . . 12
2.2 Self-assessed familiarity distributions by native group (CL1/JL1/EL1) across four target scripts (Hanzi, Kanji, Kana, English words). . . . . 14
2.3 Near–mid–far depth zones in near-eye VR displays (conceptual).[50] . . . 15
2.4 A Cross-Linguistic Comparison of Orthographic Structures. . . . . 17
2.5 Schematic illustration of the complexity measure using Japanese Kanji. . 19
2.6 Typeface Overview Diagram. (a) Comparison of area occupied by four writing systems at the same point size. (b) Examples of sans-serif versus serif font styles. (c) Sample of six fonts combinations: 2 styles (Sans, Serif) × 3 weights (Thin, Medium, Bold) across four scripts. . . . . 21

3.1 Schematic Diagram of Experimental Setups for Chinese, Japanese, and English Tests. . . . . 26
3.2 Overview of the Distance-Adjustment Task. . . . . 29
3.3 Schematic summary of observed viewing-distance effects in VR. (a) Language familiarity patterns; (b) effects of glyph complexity and font weight. 35
3.4 Schematic illustration of foveal and parafoveal processing during reading, highlighting the perceptual span [111]. . . . . 36
3.5 EMM-predicted mean viewing-distance adjustments for native Chinese (CL1), Japanese (JL1), and English (EL1) readers across all language experiments at 0.5 m (panel a) and 2.5 m (panel b). . . . . 37
3.6 EMM-predicted mean viewing-distance shifts under 10 m for different glyphs and native groups in the Japanese(a) and English(b) experiments. 38
3.7 EMM - predicted 10 m viewing - distance shifts across three experiments: simple versus complex glyphs in Japanese (a) and Chinese (b), and short(E1) versus long(E3) words in English (c), for native Chinese, Japanese, and English readers . . . . . 39
3.8 Cumulative EMM-predicted viewing-distance shifts at 2.5 m and 5 m by font weight and native-language group for simple and complex glyphs in the Japanese (a–d) and Chinese (e–h) experiments. Positive values indicate moving the panel farther; negative values indicate moving it closer. For readability, values with $|\Delta| < 0.10$ m are shown as 0 (label omitted). . 40

3.9 Cumulative EMM-predicted viewing-distance shifts at 2.5 m and 5 m in the English experiment, grouped by font family and native-language group. Positive values indicate moving the panel farther; negative values indicate moving it closer. For readability, values with $|\Delta| < 0.10$ m are shown as 0 (label omitted). . . . . 41

4.1 Schematic of the experiment scene. This figure illustrates the virtual environment in which participants viewed a text box containing either two Chinese/Japanese characters or two English words at varying distances (0.5 m, 2.5 m, 5 m, and 10 m). . . . . 46
4.2 Experiment Flow and Task Flow Diagram. This figure illustrates the overall flow of the experiment, starting with the introduction and consent signing, followed by practice sessions, the main tasks in L1 (native language) and L2 (non-native language), and concluding with participant feedback. The task flow section on the right details the specific steps participants followed during each trial, including reading the glyph prompt, selecting the matching glyph using the Quest Touch Pro controllers, and taking breaks between sets of trials. . . . . 51
4.3 Error rates for Chinese, Japanese, and English Experiments.Parts (A), (B), and (C) depict the interaction between Glyphs and Native language groups. Glyphs C1–C9, J1–J11, and E1–E7 represent increasing complexity, with odd-numbered glyphs (e.g., C3, J5, E3) being similar and even-numbered glyphs (e.g., J2, J4) dissimilar. In the Chinese experiment (A), glyphs C3, C5, and C7 show the highest error rates, with English speakers performing worse on C3 but better on C5 and C7 compared to Chinese and Japanese speakers. In the Japanese experiment (B), J9 exhibits the highest error rates, with native Japanese speakers showing consistently lower errors than others. Chinese speakers show balanced errors, while English speakers' errors increase with complexity. In the English experiment (C), overall error rates are lower, but E1, E3, and E5 show peaks, especially for Chinese and Japanese speakers, with E3 being the most error-prone glyph. . . . . 57
4.4 Error bar plots (95% confidence intervals) for Glyph-Font interactions in the English experiment (5 m) and Japanese experiment (10 m). . . . . 58
4.5 Overall invisibility rates Across Different Language Experiment Environments. The x-axis represents different language experiment environments, while the y-axis indicates the 95% confidence interval of the mean invisibility rates. Overall, the Chinese group had the highest invisibility rates, followed by the Japanese group, while the English group' s invisibility rates were nearly zero. . . . . 59

4.6 Invisibility Rates in Chinese and Japanese Experiments. Part (A) illustrates the invisibility rates for glyphs C1 to C9 in the Chinese experiment, grouped by native language (Chinese, Japanese, and English speakers).Part (B) depicts the invisibility rates for glyphs J1 to J11 in the Japanese experiment, similarly grouped by native language. Glyphs C5, C6, C7, and C9 showed higher invisibility rates in the Chinese experiment, while glyphs J7, J9, and J10 exhibited higher invisibility rates in the Japanese experiment. . . . . 62
4.7 Average Task Completion Time by Glyph and Native Language for Chinese, Japanese, and English Experiments. parts (A), (B), and (C) illustrate the interaction between Glyph and Native language groups in the Chinese, Japanese, and English experiments, respectively. In the Chinese and Japanese experiments, English native speakers generally require more time to complete tasks (parts (A) and (B)), especially for complex glyphs such as C3, C5, C7 in Chinese and J7, J9, J10 in Japanese. Under the 10 m condition, compared to 5 m, there is a notable increase in average task completion times. In the English experiment (part (C)), task completion times are similar across all native language groups, with visually similar glyphs (such as E3 and E6) requiring slightly more time than dissimilar glyphs. . . . . 67

5.1 Schematic Diagram of Experimental Setups for Chinese, Japanese, and English Tests. . . . . 78
5.2 Experiment Flow and Task Flow Diagram . . . . . 82
5.3 Overall Font Preferences of Chinese, Japanese, and English Experiments.The x-axis represents different language experiments (CN = Chinese, JP = Japanese, EN = English), and the y-axis shows the mean font scores (with 95% confidence intervals). Significant differences are marked by "*" ($*p < 0.05, **p < 0.01, ***p < 0.001$). . . . . 86
5.4 Legibility comparisons across Chinese font styles and weights at four viewing distances. Panels (a–d) present serif vs. sans-serif contrasts under each weight category, while panels (e–h) present weight contrasts (Black, Middle, Thin) within each font style. The x-axis lists font names, and the y-axis shows mean performance scores (0–2) with 95% confidence intervals. Post-hoc pairwise comparisons (EMM) are indicated by asterisks ($*p < .05$, $**$ $p < .01$, $***p < .001$). . . . . 89

5.5 Legibility comparisons across Japanese font styles and weights at four viewing distances. Panels (a–d) present serif vs. sans-serif contrasts under each weight category, while panels (e–h) present weight contrasts (Black, Middle, Thin) within each font style. The x-axis lists font names, and the y-axis shows mean performance scores (0–2) with 95% confidence intervals. Post-hoc pairwise comparisons (EMM) are indicated by asterisks ($*p < .05$, $**\ p < .01$, $***p < .001$). . . . . 91
5.6 Legibility comparisons across English font styles and weights at four viewing distances. Panels (a–d) present serif vs. sans-serif contrasts under each weight category, while panels (e–h) present weight contrasts (Black, Middle, Thin) within each font style. The x-axis lists font names, and the y-axis shows mean performance scores (0–2) with 95% confidence intervals. Post-hoc pairwise comparisons (EMM) are indicated by asterisks ($*p < .05$, $**\ p < .01$, $***p < .001$). . . . . 93
5.7 Font Performance Trends of Chinese Fonts Under Varying Glyph Complexity, Native Language Groups, and Viewing Distances. Panels (1)–(6) illustrate how font performance varies with different glyph complexities (x-axis), while Panels (7)–(12) show how performance differs across native language groups (x-axis). In all panels, the y-axis indicates the mean font performance scores (0–2 scale) with 95% confidence intervals, where "M" denotes the average score for each font. Separate post hoc analyses were conducted for each trend, with significant differences marked by "*" ($*p < 0.05, **\ p < 0.01, ***p < 0.001$). . . . . 95
5.8 Font Performance Trends of Japanese Fonts Under Varying Glyph Complexity, Native Language Groups, and Viewing Distances. Panels (1)–(6) illustrate how font performance varies with different glyph complexities (x-axis), while Panels (7)–(12) show how performance differs across native language groups (x-axis). In all panels, the y-axis indicates the mean font performance scores (0–2 scale) with 95% confidence intervals, where "M" denotes the average score for each font. Separate post hoc analyses were conducted for each trend, with significant differences marked by "*" ($*p < 0.05, **\ p < 0.01, ***p < 0.001$). . . . . 96
5.9 Font Performance Trends of English Fonts Under Varying Glyph Complexity, Native Language Groups, and Viewing Distances. Panels (1)–(6) illustrate how font performance varies with different glyph complexities (x-axis), while Panels (7)–(12) show how performance differs across native language groups (x-axis). In all panels, the y-axis indicates the mean font performance scores (0–2 scale) with 95% confidence intervals, where "M" denotes the average score for each font. Separate post hoc analyses were conducted for each trend, with significant differences marked by "*" ($*p < 0.05, **\ p < 0.01, ***p < 0.001$). . . . . 97

5.10 Heatmap Illustrating Font × Native Interactions Across Different Language Experiments (Derived from Each Font' s Winning Rate) . . . . . 98
5.11 Win Rate Trends of Fonts Across Languages and Glyph Complexity Levels 100
5.12 Comprehensive Overview of Font Preferences by Distance × Glyph, and Native Speaker . . . . . 101

6.1 Base prototype with neutral blocks and placeholder text. . . . . 108
6.2 Metro-style scene that augments the base prototype with simple information-hierarchy rules on top of distance- and complexity-aware text adaptation. 109

B.1 Power analysis for the Chinese experiment . . . . . 137
B.2 Power analysis for the Japanese experiment . . . . . 137
B.3 Power analysis for the English experiment . . . . . 138

# List of Tables


1.1 Script/language distribution of VR legibility papers (counts) . . . . . 4
1.2 Factors studied in VR legibility papers (counts) . . . . . 4
1.3 Populations/contexts in VR legibility papers (counts) . . . . . 4

2.1 Self-assessment items with anchor examples . . . . . 13
2.2 Schematic multitrait–multimethod (MTMM) structure of the thesis . . . . 22

3.1 Character–glyph classification across Chinese, Japanese, and English. CJK tiers reuse the complexity thresholds from Section 2.3.3. . . . . 28

4.1 Character Glyph Classification by Feature and Similarity Across Chinese, Japanese, and English . . . . . 49
4.2 Error Distribution Across Different Distances, Glyphs, Fonts, and Native Languages . . . . . 53
4.3 Invisible Times Distribution Across Different Distances, Glyphs, Fonts, and Native Languages . . . . . 60
4.4 Mean Task Completion Time Across Different Distances, Glyphs, Fonts, and Native Languages . . . . . 64

5.1 Character Glyph Classification by Feature Across Chinese, Japanese, and English . . . . . 80
5.2 Tests of Between-Subjects Effects by Glyph, Font, Distance, and Native Language . . . . . 87