



HAL
open science

Quantifying the Cost of Manual Navigation: A Comparison of Gesture-Based Magnification versus Direct Access Reading in Digital Layout-based Documents

Sebastián Gallardo, Aurelie Calabrese, Hui-Yin Wu, Monica Di Meo, Stéphanie Baillif, Dorian Mazauric, Pierre Kornprobst

► To cite this version:

Sebastián Gallardo, Aurelie Calabrese, Hui-Yin Wu, Monica Di Meo, Stéphanie Baillif, et al.. Quantifying the Cost of Manual Navigation: A Comparison of Gesture-Based Magnification versus Direct Access Reading in Digital Layout-based Documents. IMX 2026 - International Conference on Interactive Media Experiences, Technological University of the Shannon, Jun 2026, Athlone, Ireland. <10.1145/3788851.3805017>. <hal-05605456>

HAL Id: hal-05605456

<https://inria.hal.science/hal-05605456v1>

Submitted on 28 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Quantifying the Cost of Manual Navigation: A Comparison of Gesture-Based Magnification versus Direct Access Reading in Digital Layout-based Documents

Sebastián Gallardo*
 Hui-Yin Wu
 Dorian Mazauric
 Pierre Kornprobst
 sebastian.gallardo-diaz@inria.fr
 pierre.kornprobst@inria.fr
 Université Côte d'Azur, Inria
 Sophia-Antipolis, France

Monica Di Meo
 Stéphanie Baillif
 di-meo.m@chu-nice.fr
 baillif.s@chu-nice.fr
 CHU Pasteur
 Nice, France

Aurelie Calabrese
 aurelie.calabrese@univ-amu.fr
 Aix-Marseille Université, CNRS
 Aix-Marseille, France

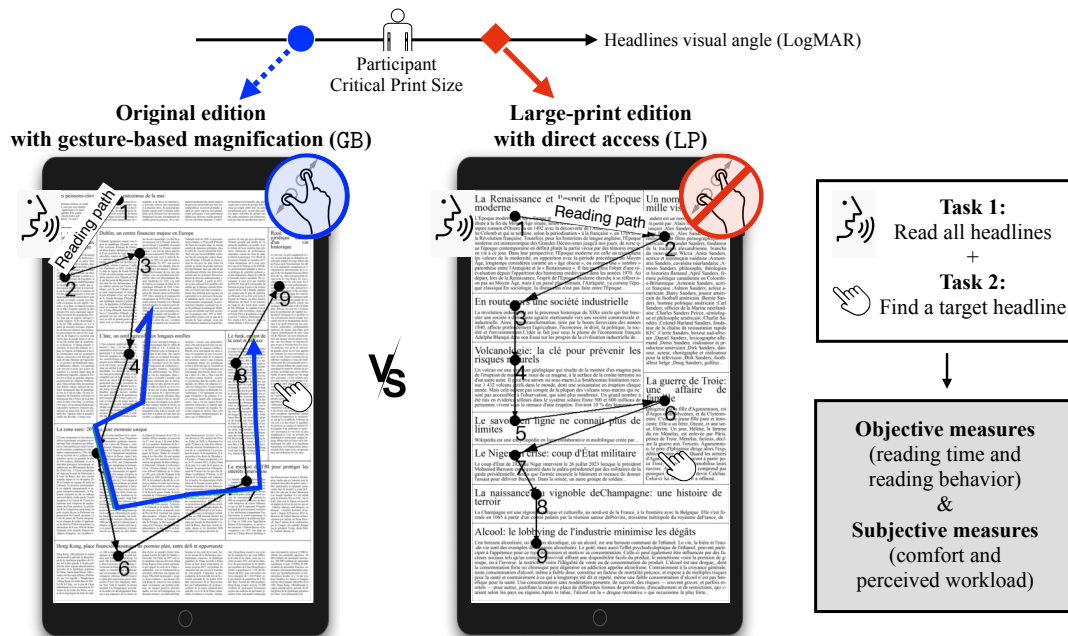


Figure 1: Experimental protocol for quantifying the behavioral and performance cost of digital newspaper consumption under visual constraints. The figure illustrates the two conditions used to evaluate the impact of manual navigation versus direct structural access when reading digital newspapers. In the gesture-based magnification condition (GB), participants read the original layout, in which headlines were rendered below each individual’s Critical Print Size (CPS), requiring pan-and-zoom magnification (high manual control, fragmented global view). In the large-print edition with direct access condition (LP), headlines were enlarged above CPS and directly legible without magnification (no manual control, direct structural awareness). Across multiple trials, participants performed two tasks—reading all headlines aloud and locating target articles—while both objective performance and subjective responses were collected.

Abstract

Understanding how diverse audiences engage with structured media is critical to ensure a consistent quality of experience. In this context, we quantify the behavioral and performance cost of manual navigation (e.g., pinch and zoom) versus direct structural access in layout-based digital documents. We specifically investigate newspaper reading when visual access to structural cues (headlines as entry points) is constrained. Participants completed two tasks—reading

*Also with Demain Un Autre Jour.



all headlines aloud and locating target articles—under two conditions: (1) original edition with gesture-based magnification (pan and zoom), which is the industry standard for digital documents, and (2) large-print edition supporting direct-access reading. We collected performance measures (success ratio and completion time), behavioral integrity through reading path analysis, alongside perceived workload and preferences (NASA-TLX). Results from linear mixed-effects models show that the large-print condition yielded not only better performance than gesture-based magnification (18% improvement in reading speed, 30% improvement in speed to locate a target), but more importantly, restored the natural reading strategy that gesture-based magnification interaction disrupts. Readers also reported lower workload and higher preference. These findings highlight the importance of developing automated methods for generating large-print editions, where layout adaptation complements font scaling to support accessibility and quality of experience.

CCS Concepts

• **Human-centered computing** → **Empirical studies in accessibility**; **Empirical studies in interaction design**; **Empirical studies in HCI**.

Keywords

Digital media consumption; Layout-based documents; Manipulable interaction; Behavioral cost; Reading behaviour; Low vision accessibility

ACM Reference Format:

Sebastián Gallardo, Hui-Yin Wu, Dorian Mazauric, Pierre Kornprobst, Monica Di Meo, Stéphanie Baillif, and Aurelie Calabrese. 2026. Quantifying the Cost of Manual Navigation: A Comparison of Gesture-Based Magnification versus Direct Access Reading in Digital Layout-based Documents. In *ACM International Conference on Interactive Media Experiences (IMX '26)*, June 09–11, 2026, Athlone, Ireland. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3788851.3805017>

1 Introduction

Digital media reading spans a continuum from linear narratives to layout-based documents such as newspapers. In layout-based media (such as newspapers), information is distributed across multiple spatially distinct blocks. Readers must not only access text at a legible size but also maintain awareness of the document's structure to ensure a consistent quality of experience. This dual requirement—legibility plus structural awareness—distinguishes layout-based reading from simpler linear reading. Entry points such as headlines, subheadings, and visual cues guide attention and shape digital newspaper consumption in ways that are unique to these documents [20, 27, 28, 60].

Effective navigation therefore hinges on visual access: the ability to perceive spatial organization and identify entry points. And so, navigation strategies break down when visual access is constrained and these cues can no longer be easily perceived. We refer to such situations as constrained visual access scenarios (CVAS). In this study, we focus on two complementary CVAS that both reduce the legibility of entry points: (1) device-related constraints, where limited screen size restricts visibility even for readers with normal

vision¹ (e.g., on small phones) and (2) reader-related constraints, where reduced visual acuity limits access to layout structure even on larger displays (readers with low vision).¹ While these scenarios have been studied in the simpler context of plain text [7], their impact on digital newspaper, which have multi-block layouts remains under-explored.

These constraints have real-world consequences that extend beyond basic legibility. Because both types of CVAS limit access to a page's structure and entry points, they hinder reading, reduce exploratory behavior, and degrade the overall quality of experience [11, 38, 39]. For readers with low vision, the same loss of visual access further restricts access to newspapers, affecting not only information acquisition but also social participation, as news media play a key role in maintaining engagement with societal discourse [23]. Ensuring that layout-based documents remain navigable under CVAS is therefore essential for equitable access to information.

A common approach to support reading under CVAS could be to simplify the content, for example, by linearizing articles into feed-like formats. While effective for legibility, these representations discard the spatial organization that fosters serendipity, engagement, and aesthetic appreciation in layout-based documents [26, 42].

When simplification is not desirable, the natural alternative is to magnify the original content. On touch devices, gesture-based magnification (namely pan-and-zoom) is the standard solution. Its adoption is well-founded: manipulable interfaces increase users' sense of control, predictability, and engagement [3, 6, 43, 47], including for readers with visual impairments [22]. However, these benefits often come at the cost of global context [13, 30]. While pan-and-zoom preserves the underlying layout, it exposes it only through successive zoomed-in fragments. Readers must zoom in to read a headline, pan to follow the article, then pan again to locate the next entry point—or zoom out to regain a sense of the page. Then, manual magnification does not just slow down the reader; it imposes a shift from the natural content consumption. This fragmentation disrupts the natural reading path, a problem well documented in maps and linear documents [10, 18, 30, 48]. Yet, prior findings cannot be assumed to generalize to layout-based documents. Newspapers feature multiple spatially distinct blocks, diverse entry points, and layered visual hierarchies [42], creating navigation demands that differ from maps or linear text.

To address these limitations while maintaining both legibility and spatial structure, another option is to consider large-print layouts (large-print edition with direct access), inspired by historical initiatives such as the New York Times Large Type editions [50]. These editions not only enlarge fonts (which might be not sufficient since layouts were not designed for this new font size [21]), but also reorganize layout blocks so that all entry points are directly legible without manual zooming. While this reorganization changes the absolute position of each element, it is specifically designed to preserve the 2D spatial paradigm of a newspaper. By maintaining a non-linear arrangement of multiple articles, the spatial qualities of the layout are preserved, ensuring the reader still engages with a 2D

¹The World Health Organization (WHO) classifies visual impairment based on best-corrected visual acuity in the better eye, ranging from "Normal vision" (20/20) to moderate (down to 20/160), severe (down to 20/400), and profound low vision (down to 20/1000).

document rather than a simplified 1D linear representation. Recent computational work [21], based on responsive design principles [4], automates this process by increasing headline size and reducing the number of visible lines per block. By magnifying entry points while preserving spatial layout, this approach acts as a direct-access reading modality, eliminating the need for gesture-based navigation, prioritizing global context at the expense of direct manipulation.

Our study quantifies the behavioral and performance cost of these two modalities: (1) whether the loss of global context undermines the advantages of manipulable interfaces, and (2) whether producing large-print edition with direct access layouts—designed to preserve global context but requiring effort or automation—is justified and for whom. To address this, we investigate how original edition with gesture-based magnification and large-print edition with direct access layouts influence behavioral integrity, reading performance, and perceived workload under constrained visual access. We hypothesize that while pan-and-zoom allows detailed inspection, it results in not only higher completion time than large-print edition with direct access (an anticipated mechanical consequence due to the elimination of manual navigation), but more importantly will disrupt the canonical reading strategy and increase the perceived workload in comparison to large-print edition with direct access. Our goal is to precisely quantify these costs through a multidimensional evaluation of performance, behavior and workload.

To test these hypotheses, we conducted a controlled user study with participants having normal vision or low-vision, comparing performance (success ratio and completion time), reading behavior (reading paths), and page structure memorization (success ratio on finding specific headline) on text-only newspaper pages under two conditions (see Fig. 1):

- (GB) Reading the original edition with gesture-based magnification, required to access and read the headlines due to constrained visual access.
- (LP) Reading the large-print edition, with enlarged text and reorganized layout blocks (based on [21]), ensuring all headlines are directly legible.

We evaluate both objective measures (success ratio, reading time, reading paths) and subjective measures (comfort and perceived workload) across two CVAS reflecting different devices and readers' visual acuity: small-screen reading by participants with normal vision, and large-screen reading by participants with low vision.

The remainder of the paper reviews related work, introduces our research questions and hypotheses, describes the experimental design, presents the results, and concludes with implications for the quality of experience in digital media consumption.

2 Related work

2.1 Digital media reading

Recent research in the interactive media experience community has begun to explore how different media-mixes, ranging from traditional articles to 360-degree video, affect user experience in immersive journalism. For instance, Bujic et al. [12] compared 360-video across devices against a 'monitor-article' condition, finding that while immersive formats may increase involvement, traditional text-based articles can offer lower levels of distraction, which is

critical in contemporary news consumption. This highlights the relevance of text-based media on digital platforms.

Within this landscape, prior research has shown that mobile devices have reshaped reading practices and have changed readers' habits [19, 46, 54]. At the same time, reading on mobile remains more demanding than on larger displays. Budiu [11] identified inherent constraints—such as limited screen size, frequent interruptions, single-window views, and unstable connectivity—that require careful design considerations to support sustained reading.

In particular, Nielsen [39] reported that comprehension of complex content on small displays is only about 48% of that on desktop monitors. This gap is explained by two main factors: mobile readers see less information at once, reducing context and thus understanding, and they need to navigate through the document, which diverts attention. Moran [38] further shows that while short, simple texts are read with similar comprehension across devices, reading speed drops significantly for longer or denser passages, highlighting the need for layout optimization and interaction strategies to improve mobile reading performance.

These challenges become even more acute with structured, layout-driven documents. These structured documents combine dense text with visual design and spatial organization to convey meaning. Due to their distinct design elements, findings derived from linear reading cannot be generalized to layout-driven documents [26, 42]. They pose distinctive challenges due to their complex structures, often organized around visual entry points [29].

It is worth noting that, in theory, web-responsive adaptation principles could address the general challenge of reading complex documents on small screens. However, preserving spatial semantics is often at odds with standard web-adaptation paradigms. Techniques such as responsive design handle small displays primarily through linear re-flow mechanisms [37, 53], collapsing multi-column grids into a single vertical stack defined by the DOM order. While this method preserves sequential reading order and eliminates bi-directional scrolling, it sacrifices the 2D spatial paradigm that defines newspapers and other layout-driven documents [16, 17].

As a result, for documents whose navigation, visual hierarchy, and interpretation rely on spatial organization [28, 41], standard responsive design—and similarly summary-based or linearized representations—cannot serve as viable alternatives: they remove visual entry points, suppress layout cues, and fundamentally transform the nature of the reading task.

This limitation motivates evaluating large-print layouts as an alternative that preserves spatial organization while restoring entry-point legibility, allowing us to examine the behavioral and performance cost of manipulable navigation (GB) compared to direct structural access (LP).

2.2 Situational visual impairments (SVIs) in mobile reading

Situational visual impairments [45] (SVIs) arise when reading is disrupted by device limitations, such as small font or display size and low resolution, or by environmental factors, including luminosity and ambient noise. Tigwell et al. [49] investigated how SVIs like glare or movement affect mobile content usability, highlighting

that designers often lack specialized tools and guidelines to address these challenges. They proposed preliminary recommendations and emphasized the need for improved design support through better resources and educational frameworks. Building on this, another study [40] specifically examined the impact of bright outdoor lighting on screen readability. Their results indicate that, while ambient brightness reduces readability, the intrinsic brightness and contrast of content play an even larger role, suggesting that design interventions should prioritize optimizing content contrast over solely compensating for glare.

In parallel, technical solutions have been proposed to mitigate SVIs. *SituFont* [58] is a dynamic font adaptation system that adjusts attributes such as size, weight, and spacing based on real-time sensor inputs and human-in-the-loop feedback. The authors validated their approach through comparative evaluations across eight simulated SVI scenarios, considering factors such as environment (indoor vs. outdoor), user mobility, and luminosity. Their findings show that *SituFont* significantly improves reading efficiency and reduces cognitive and physical workload compared to manual adjustments, highlighting the potential of adaptive font systems for supporting mobile reading under situational constraints.

While these studies provide valuable insights and technical solutions for reading under SVIs, they primarily focus on text consumption. Our work extends this line of research by investigating structured digital layout-based documents under constrained visual access scenarios (CVAS), examining the specific behavioral and performance cost of magnification strategies. In particular, we examine SVIs related to small headline font sizes and the need for magnification.

2.3 Reading with low vision

Despite the importance of inclusive design, users with disabilities have historically been underrepresented within the IMX community. A survey of 17 years of research [52] revealed that only 4.23% of papers addressed accessibility, with a specific call for more empirical studies involving people with disabilities. This gap motivates our work, which evaluates reading behavior across two distinct scenarios of visual limitation.

In this context, digital devices provide new opportunities for readers with low vision to access information through adaptable text formats and presentation options [33]. A growing body of research has examined how these readers interact with digital reading environments, including studies that simulate CVAS to better understand performance under magnification and limited visual access. In this topic, a key contribution by Atilgan et al. [7] provides a unified framework to evaluate how print size and display size impact reading speed for both small-display readers and individuals with low vision using magnified text. Their results show that limitations in font and display size can prevent some readers from maximizing reading performance.

Several studies have examined display configurations and strategies for magnified reading. Xiong et al. [57] proposed guidelines emphasizing the importance of exceeding the critical print size (CPS) and maintaining at least 13 characters per line to support fluent reading. Granquist et al. [22] asked participants with low vision to adjust text for comfortable reading, finding that most

relied on enlarging text, and to a lesser extent on reducing viewing distance. Beckmann et al. [9] studied manual navigation with CCTVs (stand-mounted video magnifiers) and reported that reading speed decreases unless the viewing window is sufficiently wide (>10 characters), consistent with the threshold identified by Xiong et al. [57]. Magnification and contrast enhancement were also shown to improve reading speed in simulated low-vision environments, with magnification generally having a more substantial effect than contrast [18].

Other work has examined navigation and magnification strategies in more detail. Bowers et al. [10] analyzed reading with optical magnifiers, highlighting challenges such as line retrace that need to be addressed in future research. Tang et al. [48] compared two types of screen magnification on modern devices: full-screen magnification (analogous to pan-and-zoom) and lens magnification (allowing users to see a larger portion of the un-magnified screen for spatial reference). They identified trade-offs in performance and usability, showing that lens mode led to more consistent and uniform mouse movements, whereas full mode caused longer and more frequent pauses. Similarly, Aguilar and Castet [1] developed a gaze-controlled system that magnifies a portion of text while maintaining a global view, which participants found more comfortable than traditional CCTVs.

Together, these studies highlight the impact of magnification on low-vision reading. However, most prior work focuses on plain, unstructured text, analyzing reading speed, errors, and navigation patterns within a fixed set of lines. In contrast, our study investigates structured layout-based documents, such as newspapers, where reading involves navigating not only text but also the spatial layout. We examine how visual access limitations affect document reading for both individuals with normal vision and those with low vision. In particular, we examine the behavioral and performance cost of gesture-based magnification (GB) compared to the direct structural access of large-print layouts (LP), quantifying how each strategy impacts the reading process.

3 Research Questions and Hypotheses

We will hereafter use the abbreviations GB (original edition, gesture-based magnification) and LP (large-print, direct access) to refer to the two experimental conditions. Our study addresses the following research questions:

- RQ1 How do the two conditions differ in terms of reading behavior on a digital newspaper page?
- RQ2 How do the two conditions affect participants' ability to form an accurate mental representation of the page layout?
- RQ3 Which condition is preferred by participants, and which one leads to a lower perceived workload?

From these research questions, we derive the following hypotheses:

- $\mathcal{H}1$ Participants in LP will show shorter reading times and faster transitions between articles than in GB.
- $\mathcal{H}2$ LP maintains a reading strategy closer to the natural reading path of the original page than GB.
- $\mathcal{H}3$ In LP, participants will locate target headlines more quickly than in GB.

$\mathcal{H}4$ In both conditions, the position of the target article will significantly influence finding time.

$\mathcal{H}5$ Participants will prefer LP over GB, reporting lower cognitive load and greater comfort.

To answer these research questions and test our hypotheses, we designed two tasks and a controlled scenario, described in the next section.

4 Methods

4.1 Overview of the study

4.1.1 General procedure. We have conducted a 2x2 split-plot design [32] study to quantify the behavioral and performance cost of digital newspaper documents under CVAS. We analyze how these individuals (readers with normal vision on small displays and readers with low vision even in larger displays) consume these digital documents.

Approval for this research was granted by Inria's Operational Committee for the Assessment of Legal and Ethical Risks.

To create controlled yet realistic reading conditions, we designed text-only newspaper pages, each composed of multiple blocks containing one headline and a body text (Sec. 4.2). This approach provides two key advantages:

- **Controlled entry points:** Headlines serve as the sole entry points, ensuring that participants' reading path relies on a textual structure that can be controlled and standardized. This also allows us to track the sequence in which participants access articles through aloud reading, a practical alternative to eye-tracking, which is difficult to implement with participants with low vision.
- **Reduced cross-trial variability:** By avoiding graphical elements (e.g., advertisements, photographs, or diagrams), which are harder to standardize, we minimize potential confounds across trials and conditions, ensuring comparability of reading behavior data.

The experiment consisted of two tasks designed as controlled proxies of real-world newspaper consumption strategies, allowing for systematic observation and quantification of reading behavior:

- **Task 1 (structural orientation): read all headlines.**
This task emulates the initial scanning phase of newspaper consumption. Participants read aloud all headlines on a page, as fast but intelligibly as possible, in any order they wished. Reading aloud allowed us to unambiguously capture the order in which headlines were read, and is both a well-established method in reading research [55, 56] and clinical assessments (e.g., MNREAD). This task addressed $\mathcal{H}1$ and $\mathcal{H}2$. There was no time limit. Participants decided themselves when they had finished, even if some headlines were missed.
- **Task 2 (targeted information retrieval): find a target headline.**
This task tests the user's mental representation of the page and their ability to retrieve specific information. Participants were asked to locate a specific headline containing a given keyword, to probe their mental representation of page structure. This task tested the hypothesis $\mathcal{H}3$. To examine $\mathcal{H}4$,

we divided the page into three vertical zones (top, middle, bottom) and selected two targets per zone. The order was randomized for each participant, and the zone division was never disclosed. Each search trial was limited to one minute.

Both tasks were performed under GB and LP. A critical design feature of our protocol was to ensure that participants had to rely on pan-and-zoom in GB but could comfortably read headlines in LP. To achieve this, we estimated each participant's Critical Print Size (CPS) [34], defined as the smallest print size allowing readers to maintain maximum reading speed, and expressed in logMAR units (logarithm of the Minimum Angle of Resolution), which quantify text size in terms of visual angle rather than physical dimensions, which is crucial because it jointly depends on both print size in millimeters and viewing distance. Headlines in GB are below each participant's CPS, forcing gesture-based magnification, while in LP they were manually set above CPS, enabling comfortable direct reading.

As a note, we also considered integrating eye-tracking to capture complementary gaze data. However, we opted against it due to specific challenges for this work. First, in this context, fixations mix cognitive inspection with smooth pursuit driven by viewport motion, and gaze positions must be continuously remapped to a moving document coordinate system, which limits interpretability [25, 31, 51]. Second, there is a limitation to obtain a precise calibration with participants with low vision, since precise calibration requires an intact fovea which is usually not the case [25, 56]. Although technically feasible thanks to workarounds that mitigate the underlying issues, these constraints would have reduced the validity of the resulting metrics for our specific goal, which was to characterize reading strategies at a macroscopic level, not fine-grained oculomotor behavior.

Following these principles, the study protocol unfolded in five steps, summarized in Fig. 2.

- (1) Participants first completed a short questionnaire about gender, age group, and reading habits (general and news-specific).
- (2) We ran a standard MNREAD test [34] to estimate each participant's Critical Print Size (CPS).
- (3) Based on the CPS, we defined how the two conditions would be instantiated (see Sec. 4.3), including the choice of device (small-screen phone for participants with normal or corrected-to-normal vision, large-screen tablet for participants with low vision) and the viewing distance to be maintained during the experiment.
- (4) After a detailed explanation of the protocol, participants proceeded to the main experiment. Each block corresponded to one condition, with the order counterbalanced across participants.
 - To familiarize themselves with the tasks, participants first completed as many practice trials as needed. We used three fixed layouts (same order for everyone), and cycled through them if more than three trials were required. In practice, one or two trials were sufficient.
 - Participants then completed M test trials per condition. Number of trials M depending on the participant scenario (see Sections 4.1.2 and 4.1.3). Layouts were randomized,

and content was randomly assigned to conditions (see Section 4.2)

- After each block, we measured perceived workload using the NASA-TLX questionnaire [24], using a 7-point likert scale
- (5) Finally, participants indicated their overall preference between GB and LP (or equal preference).

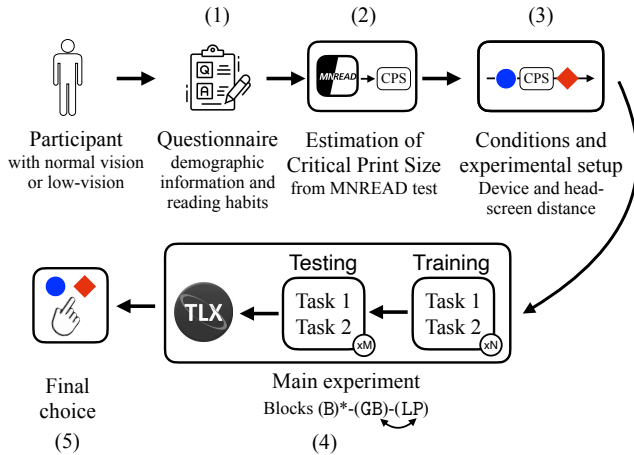


Figure 2: Overview of the study. This figure illustrates the workflow of our experimental protocol, organized into five main steps: (1) questionnaires, (2) estimation of Critical Print Size (CPS), (3) conditions and experimental setup, (4) the main experiment, and (5) a final preference choice. The main experiment consisted of successive blocks in the two conditions (GB and LP). Some adjustments were made depending on the participant scenario. Each block included a training phase followed by test phases, where Tasks 1 and 2 were repeated across trials, and ended with the NASA-TLX questionnaire. The overall protocol lasted less than an hour, with the main experiment capped at 30 minutes to avoid excessive fatigue for participants with low vision.

This protocol served as the common foundation of our study. However, some adjustments were necessary depending on the participant’s scenario.

4.1.2 Protocol adjustments: participants with normal vision. First, after pilot studies to study participants’ fatigue, we decided to fix $M = 6$ testing trials for participants with normal vision.

Second, we included an additional baseline condition (B), applied before GB and LP, with the same number of trials (N for practice and $M = 6$ for testing).

B simulates the experience of reading a physical newspaper, corresponding to unconstrained reading. In this condition, the original edition was displayed on a large-screen tablet, with no magnification gestures, as headlines were already comfortably legible (approximately 0.3 logMAR at 40 cm). Note that this baseline was not feasible for participants with low vision, since the original edition would remain unreadable without magnification, even on a tablet. Data from B were used to define a natural reading path, which we

call reference paths, and these were compared to the paths observed in GB and LP when testing hypothesis $\mathcal{H}2$.

4.1.3 Protocol adjustments: participants with low vision. For participants with low vision, pilot sessions revealed that 6 testing trials were overly demanding for this group: several participants suddenly experienced acute visual fatigue, making it impossible to continue and leading to early interruptions. As a consequence, this protocol design risked unbalancing the data, with many completing only a single condition. To address this, we reduced and reorganized the trials and blocks (changes in Fig 2 (4) and (5)):

- Participants first completed two trials per condition, each followed by a NASA-TLX questionnaire, to ensure that subjective workload was captured before cumulative fatigue could bias responses.
- The order of layouts was randomized but kept identical across GB and LP to support fair comparisons.
- To maximize usable data, participants continued with the remaining four trials per condition in an alternating sequence (one GB, one LP, and so on). Sessions ended as soon as all trials were completed, the participant indicated they were too tired to continue, or the 30-minute time limit was reached.
- Finally, participants reported their overall preference between GB and LP, or both equally. Because cumulative fatigue strongly influenced judgments, we also administered a comparative NASA-TLX at the end of the session to balance condition evaluations.

Having outlined the overall procedure, the following subsections expand on its key components: the creation of newspaper page material, the conditions established based on CPS, the population, and the data collected for subsequent analysis.

4.2 Newspaper pages material

4.2.1 Original edition. Each newspaper page combined two components: (1) a *layout file*, defining the geometric arrangement of articles on the page, and (2) a *content file*, providing the textual material for each article (headline, body text, and target font sizes). Because our study focuses on headlines, body text was generated only to fill space within article blocks and was not analyzed.

In total, we generated 18 newspaper pages, each containing between 8 and 11 articles (see Fig. 3 (a)). Below, we detail the construction of layouts, headline sizes, and content.

Font size. Headline size was fixed to correspond to a visual angle of 0 logMAR when displayed on the phone at a viewing distance of 40 cm (Fig. 3 (a)); see Sec. 4.3). This 0 logMAR size was intentionally chosen to challenge normally-sighted participants, as it requires zooming for comfortable reading [15]. Body text size was set heuristically to half the headline size.

Layouts. Layouts were created manually from annotated front pages of the *New York Times*. We produced nine layout files ($\mathcal{L} = \{l_i\}_{i=1,\dots,9}$), each specifying the number of articles, their position, and their dimensions (width and height in mm). Based on these dimensions and the font size, we extrapolated the required article length in characters and lines.

More precisely, concerning headlines, to match the New York Times design, we first analyzed headline lengths relative to article

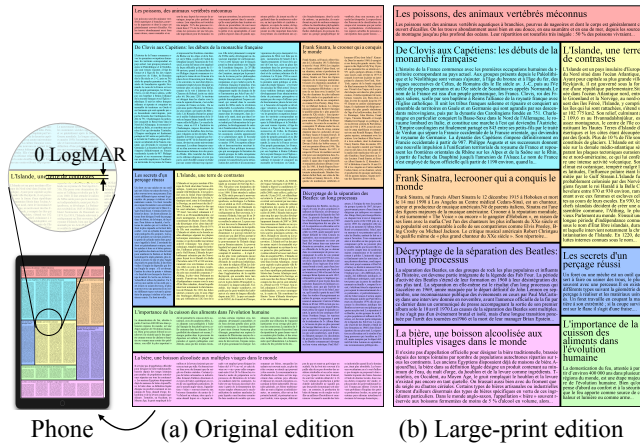


Figure 3: Examples of newspaper pages: (a) original edition used in condition GB, and (b) corresponding large-print edition used in condition LP, automatically generated using the re-laying-out method from Gallardo et al. [21] with a $\times 2$ magnification. Colors indicate matching articles across the two versions (i.e., occupying the same area on the page). The textual content differs between (a) and (b) to avoid repetition effects, but was balanced in headline length. On the left-hand side, we also illustrate how the original edition would look in a phone, which is the display used to define the headline size in LogMAR.

width (number of columns). This showed that headlines in narrower articles (spanning fewer than two columns) typically used up to three lines, while those in wider articles (three or more columns) almost always fit on a single line. We reproduced this distribution when generating our own headlines (see example in Fig. 3 (a)).

Content. Given article length specifications, we created three distinct content files for each layout file, yielding a total of 27 sets of article content $C = \{c_{ij}\}_{i=1,\dots,9, j=1,2,3}$. This ensured that participants were not exposed to the same material across conditions (B, GB, LP), avoiding memory bias. Content files were created in R [44] by selecting the article’s text from the web. For each article, a matching headline was then generated automatically with MistralAI [2]. We utilized this LLM-based approach to generate controlled media stimuli with uniform linguistic difficulty. This ensured that the behavioral integrity of the reading paths was measured against standardized content, eliminating potential engagement biases or emotional confounding that could arise from prior knowledge of real-world news events.

As a starter, a list of 364 key words (e.g., dog, Paris, etc.) was manually written to serve as article topics and sorted within 30 general themes (e.g., animals, cities, etc.). For each keyword, the R [44] *getwiki* library was used to search the web and extract the first paragraph of the top 20 Wikipedia and Vikidia pages (in French) matching the search result. This resulted in a list of 11,565 articles. Using a custom-based routine in R, this list was searched to assign the optimal candidate articles to each layout file, based on the

required character length, checking that the same theme was only chosen once within a single page to avoid redundant topics.

Last, MistralAI was prompted from R to generate a list of 10 titles per article automatically. The prompt sent to MistralAI was as follows: "Create 10 possible headlines in the style of a journal article, in French, with approximately n characters (space included), about the following article t ", n being the number of characters given by the layout files and t being the text of the article. Each output was manually inspected to select the optimal headline.

4.2.2 Large-print edition. For each original edition, we generated a large-print counterpart by applying the automatic re-laying-out method from Gallardo et al. [21]. This approach addresses the limitations of simple font scaling (which can cause text overflows and break entry points because original layouts were not designed for increased sizes [21]). This method takes an original edition and a font size magnification factor to apply, and uses an evolutionary algorithm to reorganize layouts in a way that optimizes the number of lines of each headline and the aesthetic quality of the layout e.g., alignment of articles, visual balance, etc. It is crucial to note that the goal of this adaptation is structural integrity over the preservation of absolute coordinates. While the articles are re-positioned to accommodate the larger font size, the 2D layout topology is maintained.

We feed this method with a layout l_i and one content set $(c_{i,1})$, together with a magnification factor of $\times 2$, which comfortably covers the critical print size spectrum across both participant groups (see Sec. 4.3). Only $c_{i,1}$ was used for generating these layouts (since $c_{i,2}$ and $c_{i,3}$ are equivalent in terms of length). The output is a new large-print layout $\mathcal{L}' = \{l'_i\}_{i=1,\dots,9}$ (Fig. 3 (b)). The large-print edition was obtained by combining the large-print layout with one of the three equivalent content files for that page. Note that in the large-print edition, a larger font size inevitably prevents the full body text of each article from being displayed in the overview. This is an inherent feature of these large-print digital editions, which, like existing kiosk-style news applications, are designed primarily to support scanning and discovery of articles. Readers would use the overview to identify entry points (e.g., headlines), and then would access the full content in a dedicated single-page reading mode.

4.3 Conditions setup based on CPS

To recap, our goal is to design an experiment where condition GB forces participants to use gesture-based magnification, as the original edition’s headlines are deliberately too small to read comfortably, while condition LP allows participants to read the large-print edition without any magnification. Importantly, for each participant, both conditions are presented on the same device and at the same head-to-screen distance. The procedure to achieve this setup is detailed in the following steps.

4.3.1 Estimation of the CPS. To assess participants’ CPS, we used a standardized reading test based on the MNREAD chart [35, 36]. The MNREAD is a continuous-text reading acuity chart consisting of short sentences presented at progressively smaller print sizes. Reading speed is plotted as a function of print size, producing the characteristic MNREAD curve: a plateau of constant reading speed

at larger sizes, followed by a sharp decline once print falls below a critical threshold. The CPS is defined as the smallest print size that still supports maximum reading speed, i.e., the lower bound of the plateau [34].

4.3.2 Headline size in LogMAR for both conditions. The visual angle of headlines depends on two factors: the device’s screen size and the viewing distance. Given a participant’s CPS, the goal is therefore to select the appropriate device and distance so that the headline size of the original edition in GB is smaller than the CPS, while the headline size of the large-print edition in LP is larger than the CPS. When multiple configurations are possible, we systematically choose the one that yields headline sizes closest to the CPS. This principle is illustrated in Fig. 4 (a).

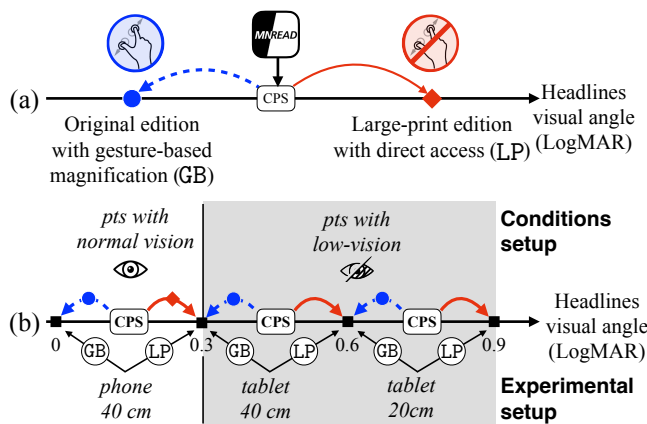


Figure 4: Experimental condition setup based on CPS. (a) General principle: headline visual angles in both conditions are selected around the participant’s CPS value obtained from an MNREAD test. (b) Device and viewing distance configuration depending on the scenario: for participants with normal vision ($0 \leq \text{CPS} \leq 0.3$), the experiment was conducted on a phone at a fixed distance of 40 cm; for participants with low vision ($0.3 \leq \text{CPS} \leq 0.9$), the experiment was conducted on a tablet, with the viewing distance adjusted according to the CPS value.

A custom-designed app, developed using React Native framework, was deployed for the application of this protocol. Regarding this, we use two different Android devices:

- **Phone:** Google Pixel 7, featuring a 7.3-inch screen with a maximum resolution of 1080×2400 pixels.
- **Tablet:** Samsung Galaxy Tab S9 FE+, featuring a 12.4-inch screen with a maximum resolution of 1600×2560 pixels.

These specifications, combined with adjustments of the viewing distance, allowed us to cover participants with CPS values ranging from 0 to 0.9. The actual configuration depended on each participant’s CPS, as illustrated in Fig. 4 (b). For participants with normal vision, we used the phone at a fixed viewing distance of 40 cm. For participants with low vision, we used the tablet, with the viewing distance adjusted according to their CPS.

4.4 Population

We recruited 24 participants: 19 participants with normal or corrected-to-normal vision and 5 participants with low vision. The sample size meets SIGCHI standards for behavioral studies [14]. The number of participants with low vision ($n = 5$) is considered acceptable when conducting research with users with disabilities [32].

Participants with normal vision were mainly recruited through an open survey, while participants with low vision were recruited from the CHU Pasteur ophthalmology department. All participants took part voluntarily. Demographic details for both groups are provided in Table 1.

ID	Age group	Gender	CPS (LogMAR)	Reading frequency	Diagnosis
1	20-35	Male	0.15	Daily	NA
2	20-35	Female	0.1	Rarely	NA
3	20-35	Female	0.1	Never	NA
4	20-35	Male	0.15	Daily	NA
5	20-35	Male	0.1	Daily	NA
7	20-35	Male	0.15	Daily	NA
8	20-35	Female	0.2	Daily	NA
9	36-65	Male	0.15	Daily	NA
10	20-35	Male	0.1	Weekly	NA
11	20-35	Male	0.2	Rarely	NA
12	20-35	Male	0.1	Rarely	NA
13	36-65	Female	0.1	Daily	NA
14	36-65	Male	0.1	Daily	NA
15	20-35	Female	0.1	Weekly	NA
16	20-35	Male	0.1	Weekly	NA
17	20-35	Male	0.08	Weekly	NA
18	66-80	Male	0.1	Daily	NA
19	36-65	Female	0.2	Daily	NA
20	66+	Female	0.15	Daily	NA
21	80+	Male	0.8	Daily	Glaucoma
22	80+	Male	0.5	Daily	Glaucoma
23	80+	Male	0.45	Weekly	Glaucoma
24	66-80	Female	0.9	Never	AMD
25	66-80	Female	0.9	Daily	AMD

Table 1: Participant demographics, CPS, and newspaper reading frequency. The last column lists diagnoses for participants with low vision, with corresponding rows shaded in light gray. Skipped ID corresponds to a dropped participant because of unusable data.

4.5 Data collected and measures

For each participant, we recorded the following data during the experiment to provide a multidimensional assessment of the quality of experience:

- **Pre-experiment questionnaire responses:** Age group, gender (optional), and news reading habits.
- **Critical Print Size:** CPS in logMAR units, obtained from the MNREAD reading test.

- **Task timing:** Total duration of each task in each trial (reading headlines, finding target articles), measured in seconds by the application.
- **Task success:** For each trial, the percentage of read articles in task 1, and also a boolean indicator for whether the participant correctly found the target article in task 2. Both of these for each trial.
- **Voice recordings:** Audio of participants reading aloud headlines, enabling reconstruction of reading order.
- **Screen interactions:** For GB trials, detailed logs of gestures, including zoom level, pan position, timestamps, and sequence, allowing reconstruction of reading paths.
- **Subjective workload:** Responses to the NASA-TLX questionnaire were proposed after each condition.
- **Overall condition preference:** Participant choice between GB and LP, or both equally preferred.

For participants with low vision, time limits and adapted trial sequences were applied as described in Sec. 4.1, and all data were recorded consistently across conditions to enable fair comparisons.

Note that, except for voice recordings, no personal data was collected. Voice recordings were stored temporarily on an encrypted laptop and processed locally using a semi-automatic speech-to-text service to extract reading order and timestamps. Once transcription was completed, the recordings were permanently deleted.

5 Results

Our analysis tested hypotheses $\mathcal{H}1$ - $\mathcal{H}5$. We first introduce the statistical model, then report results for Task 1 and Task 2, and finally analyze subjective workload and user preference using the NASA-TLX questionnaire together with overall preference.

5.1 Statistical model

We analyzed the data using a linear mixed-effects model with Participant ID as a random effect. The fixed effects are detailed in Table 2. We ran an ANOVA to test for significance using the F-statistic, setting the p-value threshold at $p < 0.05$. All reported values are model estimates (i.e., estimated marginal means) rather than raw data.

The following sections report exact p-values and effect sizes for significant main effects and interactions. Complete model statistics—including F-values, degrees of freedom, and detailed pairwise comparisons—are provided in the supplementary tables in the Appendix. We do not report main effects and interactions that were not significant.

5.2 Task 1 analysis

We first transcribed participants’ audio recordings using WhisperX [8], reconstructing a task timeline for each participant. Each timeline was segmented into reading (R) and transition (T) periods (see Fig. 5). From this timeline, we derived four metrics to test hypotheses $\mathcal{H}1$ and $\mathcal{H}2$:

- (1) **Success ratio:** The proportion of correctly read articles out of the total (e.g., $5/6=83.3\%$ in Fig. 5). Note that a headline was considered correctly read if its transcription was phonetically equivalent to the source text, which we verified using the *phonetic-fr* Python library.

#	Factor (#Levels)	Levels	Interactions
1	Condition (2)	GB, LP	With #2, #3, #4, #5, #8
2	NewspaperID (6)	1,...,6	With #1, #3
3	CVAS (2)	Small Screen, Low Vision	With #1, #2
4	Reading frequency (4)	Daily, Weekly, Rarely, Never	With #1
5	Target headline zone (3)	Top, Middle, Bottom	With #1
6	Trial index (6)	1, . . . , 6	Main effect only
7	Content version (3)	1, 2, 3	Main effect only
8	Condition order (2)	GB first, LP first	With #1

Table 2: Fixed effects of the Linear Mixed-Effects Model (LMM) used in this study. Factor of interests are shown with a white background; secondary or control factors in gray.

- (2) **Completion time:** The total task duration, from stimulus onset to completion ($t_e - t_0$ in Fig. 5).
- (3) **Reading path:** The sequence in which articles were read (e.g., 1-3-2-6-5 in Fig. 5).
- (4) **Reading and transition time:** The durations of individual reading and transition periods.

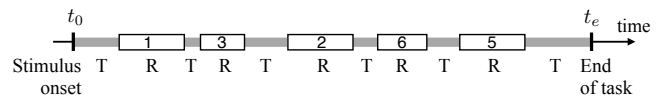


Figure 5: Task 1 timeline. After stimulus onset (time t_0 , when the newspaper page is displayed), participants alternate between transition periods for locating the next headline (T) and reading periods (R, with white boxes indicating the headline index). The task ends at time t_e , when the participant indicates they have read all headlines—even if some were missed, as illustrated in this example where headline 4 was skipped.

5.2.1 *Success ratio.* Our analysis revealed significant effects for three factors:

- **NewspaperID:** Although significant differences existed between newspapers ($p < .001, \eta_p^2 = 0.12$), performance was consistently high, with the lowest-performing newspaper still achieving a 93.4% success ratio.
- **CVAS:** Participants with normal vision achieved a higher success ratio (100%) than participants with low vision (93.4%), consistent with the reduced vision in the latter group ($p < .001, \eta_p^2 = 0.55$).
- **Content version:** A significant difference was found between Content #1 (96.2%) and #3 (98.6%), but both yielded high success ratios ($p = .025, \eta_p^2 = 0.03$).
- **A significant interaction between CVAS and NewspaperID** ($p = .002, \eta_p^2 = 0.08$) indicated that the performance gap due

to vision loss was layout-dependent, something not present in Layouts 16 and 17.

Most importantly, we found no significant effect of Condition, indicating that neither GB nor LP affected the participants' ability to read headlines correctly.

5.2.2 Completion time. Our analysis revealed significant effects for all our factors of interest:

- **Condition:** The main result is that, as anticipated due to the reduction in navigational overhead, LP had 37% lower completion time than GB (GB: 71.7 s; LP: 45.2 s) ($p < .001$, $\eta_p^2 = 0.42$). An interaction with CVAS ($p < .001$, $\eta_p^2 = 0.41$) showed that this improvement was larger for participants with low vision, with a 42% reduction (GB: 114.7 s; LP: 66.3 s) compared to a reduction of 16% for participants with normal vision (GB: 28.6 s; LP: 24.1 s). This finding supports Hypothesis $\mathcal{H}1$.
- **CVAS:** Individuals with normal vision (average of 26.3 s) had 70% lower completion times than readers with low vision (average of 90.5 s), consistent with the reduced vision in the latter group ($p < .001$, $\eta_p^2 = 0.73$).
- **NewspaperID:** Layout #17, featuring a feed-like format (i.e., articles displayed as a stack of rectangles, as commonly seen in social media feeds) had significantly shorter completion time (avg. 48.9 s) than the others (avg. 60.3 s) ($p < .001$, $\eta_p^2 = 0.19$). While effective, we consider this a special case with limited generalization (further discussed in Sec. 5.4).
- **Content version:** A significant difference ($p = .04$, $\eta_p^2 = 0.03$) was found between version #1 (avg. 55.9 s) and #3 (avg. 61 s), likely due to subtle variations in content complexity that influenced processing time or even induce more mistakes.

5.2.3 Reading path. To analyze reading paths, we first identified a reference reading strategy in B trials. Visual inspection initially suggested a left-to-right, top-to-bottom reading pattern in B trials. We then quantified the similarity between each path and this candidate pattern using a similarity measure² and an ANOVA confirmed no significant deviation from it. Based on this validation, we define the *reference path* as the theoretical left-to-right, top-to-bottom reading order implied by the page layout, independent of any experimental data.

Using the reference paths, we computed the similarity of GB and LP reading paths against reference paths induced over the corresponding layouts. To enable sequence comparison, we indexed the articles in each condition (GB and LP) according to the reference strategy (e.g., Article #1 is the top-leftmost). Consequently, a path adhering to the reference strategy results in the canonical sequence $[1, 2, \dots, n]$, while any deviation will be captured by the similarity measure.

It is important to note that we do not treat deviation from this reference path as a performance failure or a "negative" outcome. Rather, we interpret such deviations as behavioral evidence of a strategic shift in reading, quantifying the extent to which each condition disrupts the structural perception of the 2D layout.

²Given two paths p_1 and p_2 , we defined their similarity as $Sim(p_1, p_2) = 100 * (1.0 - levenshtein(p_1, p_2) / \max(\text{length}(p_1), \text{length}(p_2)))$, using the Levenshtein distance [59]. The similarity $Sim(p_1, p_2)$ ranges from 0, indicating completely different paths, to 100, indicating $p_1 = p_2$.

Our analysis revealed three significant factors:

- **Condition:** The main effect observed was that reading paths in LP were significantly more similar to the reference than in GB (GB: 69.3%; LP: 87.5%) ($p < .001$, $\eta_p^2 = 0.14$). This effect was present for both participants with normal vision (GB: 72.3%; LP: 88.9%) and for participants with low vision (GB: 66.2%; LP: 86.1%). The lower similarity in GB reveals a measurable behavioral cost: while manual interaction provides control, the navigational overhead required to manage successive zoomed-in fragments disrupts the reader's ability to maintain their natural exploration strategy. This effect is visible in Fig. 7, where participants exhibited reading paths guided by their screen interactions (i.e., panning movements), and in particular, participants with low vision required substantially more gestures, reflecting the greater effort. This observation supports hypothesis $\mathcal{H}2$.
- **NewspaperID:** Layout influenced the reading path ($p < .001$, $\eta_p^2 = 0.13$). While LP generally produced paths more similar to the reference, a significant interaction showed that ($p < .001$, $\eta_p^2 = 0.18$) layout #19 (Fig. 3) was a notable exception, where GB scored slightly higher (GB: 72.2%; LP: 68.6%). This may be due to its unique vertical organization. However, the main effect of LP superiority held across all other layouts.
- **Reading frequency:** We found an interaction between condition and reading habits ($p < .001$, $\eta_p^2 = 0.08$). In GB, frequent readers (daily or weekly) adhered more closely to the reference path than infrequent readers (who averaged only 51.4% similarity). In contrast, LP paths were highly similar to the reference (over 83%) across all reading frequencies, demonstrating their robustness.

5.2.4 Participant-level performance. To visualize individual performance, Fig. 8 plots reading path similarity against completion time for each participant, separated by CVAS to account for different data ranges. The absolute performance plots (a,c) confirm that LP resulted in shorter times and paths more similar to the reference compared to GB. The relative improvement plots (b, d) make these performance gains explicit. The plots also highlight that the improvements in both metrics were substantially larger for participants with low vision than for participants with normal vision.

5.2.5 Reading and transition times. Our analysis of reading time revealed four significant factors:

- **Condition:** There is a significant difference ($p = .01$, $\eta_p^2 = 0.05$) between LP and GB (GB: 4.18 s; LP: 3.80 s). However, due to the small effect size, we cannot confirm that reading time is the primary factor influencing the difference in completion time. Additionally, a significant effect in the interaction with NewspaperID ($p < .001$, $\eta_p^2 = 0.11$) was found. Post-hoc analysis indicated that the reading speed advantage of LP was isolated to specific layouts (Layouts #13 and #18, both $p < .001$), while no significant differences were found in the others. This indicates that the benefit on reading time is not universal, but highly dependent on specific structural arrangements.

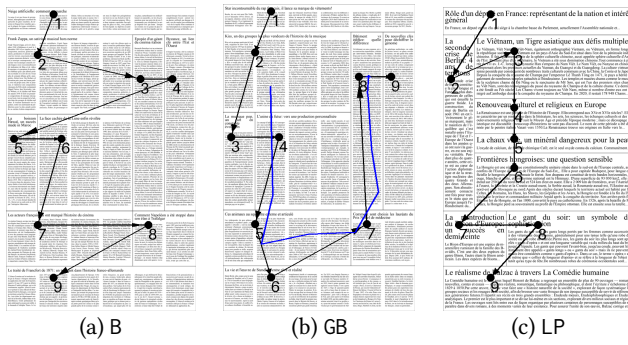


Figure 6: Examples of reading path corresponding to the same participant and layout in all conditions. (a) B path, re-assembling always the reference path (top-to-bottom, left-to-right). (b) GB path, with the screen interactions i.e., panning movements (in blue). Zoom level is not shown to simplify the figure and emphasize gesture patterns. (c) LP path.

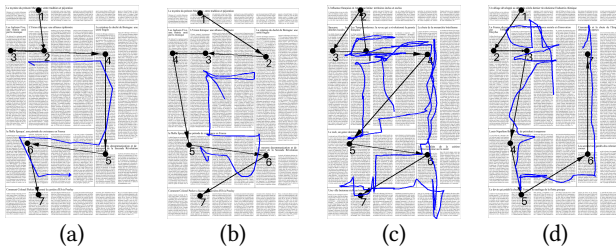


Figure 7: Four examples of GB reading paths (black) with screen interactions (panning movements in blue) for the same layout (#17). In GB, reading follows panning gestures, which do not necessarily align with the reference path. (a,b) show two different participants with normal vision. (c,d) show two participants with low vision, exhibiting similar behavior but with considerably more gestures, indicating that the task is more demanding for this group.

- CVAS: Participants with low vision had lower reading times (avg. 5.29 s) than participants with normal vision (avg. 2.68 s) ($p < .001, \eta_p^2 = 0.46$).
- Trial index: We found a significant but minor fatigue effect, with participants' reading times increasing in later trials, indicated by a positive slope (0.05) for this factor in our model ($p = .006, \eta_p^2 = 0.04$).
- Content version: A minor but significant difference ($p = .02, \eta_p^2 = 0.03$) occurred between version #1 (avg. 3.84 s) and version #2 (avg. 4.0 s), likely due to subtle variations in content complexity.

Our analysis of transition time revealed four significant factors:

- Condition: There is a significant difference ($p < .001, \eta_p^2 = 0.29$) between GB and LP (GB: 4.13 s; LP: 1.83 s). This effect size confirms that the transition time is the primary factor

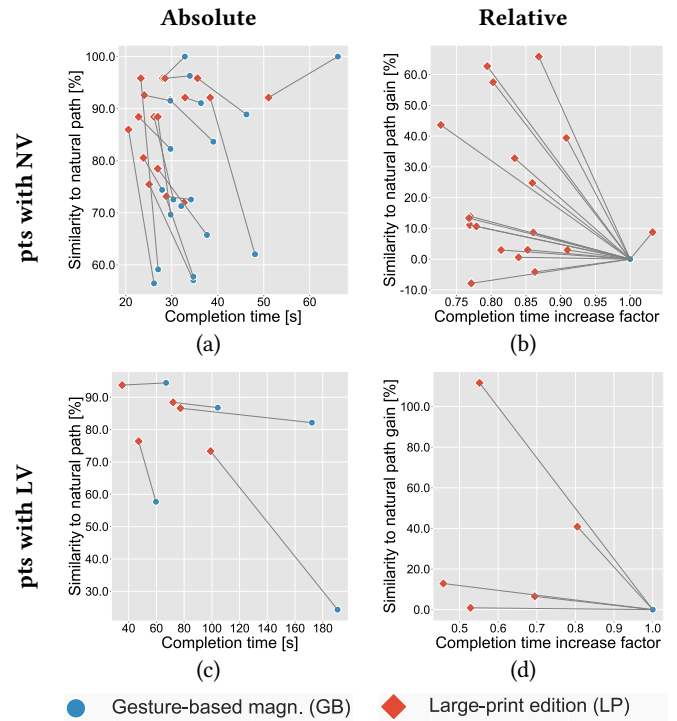


Figure 8: Average similarity to reference path vs average completion time per participant. Absolute (left column) and relative (right column) performance plots for participants with normal vision (top row) and low vision (bottom row). Absolute plots (a,c) show the average similarity to the reference path as a function of average completion time per participant, for both GB (blue circles) and LP (red diamonds). Relative plots (b,d) show the gains in similarity and the change in completion time when using LP (red diamonds) relative to GB (blue circles).

influencing completion time. However, the significant interaction with CVAS ($p < .001, \eta_p^2 = 0.35$), and the corresponding post-hoc analysis showed that this significant effect occurred only for participants with low vision ($p < .001$), who had 57% lower transition time in LP (GB: 8.0 s; LP: 3.47 s). This reduction was an expected mechanical consequence of eliminating time-consuming manual interaction adjustments and facilitated direct access of the layout, allowing for immediate transition between headlines. In contrast, no significant difference was found for participants with normal vision ($p = 0.75$), likely due to their high efficiency and device familiarity in both conditions. This result supports $H1$.

- CVAS: As expected, there was a large performance gap between groups ($p < .001, \eta_p^2 = 0.80$), with participants with normal vision (avg. 0.214 s) transitioning with a 96% shorter time than participants with low vision (avg. 5.74 s).
- NewspaperID: There is a significant difference ($p < .001, \eta_p^2 = 0.15$), driven by Layout #13 which was significantly slower in

transition time (avg. 4.05 s) than the others (avg. 2.76 s), likely due to a unique layout with a central article surrounded by the others, which might make the reading task more difficult. In any case, the main effect of LP superiority held in this layout, as well as in all the others.

- **Trial Index:** Unlike reading time, we found a significant (but minor) reduction in transition time across trials, indicated by a negative slope (-0.09) for this factor in our model ($p = .037$, $\eta_p^2 = 0.02$). This suggests a learning effect regarding the experimental setup: participants became more proficient as the experiment progressed, even as reading fatigue set in.

Figure 9 visualizes the average reading and transition times for each participant, separated by CVAS. The plots confirm that both reading times (a, c) and transition times (b, d) were shorter with LP than with GB.

Moreover, the transition time plots (b, d) reveal that the difference is primarily driven by the first and last transitions. This reflects the initial setup time required in GB for participants to adjust the zoom and pan to the first article. This effect was particularly pronounced for participants with low vision, who were less familiar with touch gestures. These graphs underscore how layouts that eliminate manual interaction can significantly improve reading performance for users in CVAS.

5.3 Task 2 analysis

Regarding Task 2, we derived two metrics to test hypotheses $\mathcal{H}3$ and $\mathcal{H}4$:

- (1) **Success ratio:** for this task, the success ratio was a binary measure (1 = target headline found, 0 = not found).
- (2) **Completion time:** The total task duration, from stimulus onset to completion.

5.3.1 Success ratio. Our analysis revealed a significant effect of NewspaperID ($p = 0.001$, $\eta_p^2 = 0.09$), and an interaction between NewspaperID and CVAS ($p = .006$, $\eta_p^2 = 0.07$) related to one specific newspaper layout (#13). This was traced to a single participant with low vision who, likely due to fatigue on his final trials, failed to find the target in either condition. Given the small sample size, this single failure disproportionately skewed the result for that specific case (48.9% success ratio).

Apart from this statistical artifact, no other factors had a significant effect on the success ratio. Overall, participants performed very well, achieving a success rate of over 90% in both the GB and LP conditions.

5.3.2 Completion time. We extracted completion time for each trial, measured in seconds. Our analysis identified two significant factors:

- **Condition:** There is a significant difference between conditions ($p < .001$, $\eta_p^2 = 0.12$). Overall, participants had 50% shorter completion time with LP than with GB (GB: 18.7 s; LP: 9.2 s). This effect was observed across both CVAS as shown by the significance of the interaction ($p < .001$, $\eta_p^2 = 0.07$), with a 51% improvement for participants with normal vision (GB: 6.91 s; LP: 3.36 s) and a 50% improvement for participants with low vision (GB: 30.37 s; LP: 15.03 s). We attribute this to the navigational cost of manual interaction:

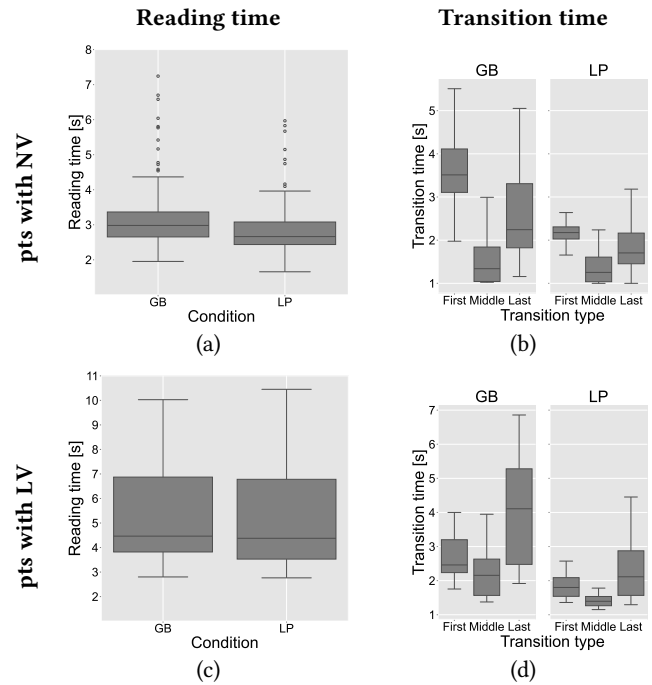


Figure 9: Reading and transition time per condition (Task 1). Reading time (left column) and transition time (right column) boxplots for participants with normal vision (top row) and with low vision (bottom row). (a, c) shows reading time boxplots. (b, d) shows transition time boxplot, separated by transition type: first transition (from stimulus onset to first headline), average of middle transitions (between headlines) and last transition (from last article to end of task). Outliers were removed to improve visualization

in GB, the effort required to manage the interaction distracts from the structural retrieval task. In contrast, LP allows for direct perception of the layout, reducing the time spent on interface management. This result provides strong support for $\mathcal{H}3$.

- **CVAS:** Participants with normal vision located the target considerably faster (avg. 5.14 s) than participants with low vision (avg. 22.7 s) ($p < .001$, $\eta_p^2 = 0.71$).

We can observe these results in Fig. 10.

5.3.3 Headline zone in Task 2. Finally, we explored whether the target headline's location ("zone") influenced task completion time. The analysis revealed no statistically significant effect ($p = .067$, $\eta_p^2 = 0.03$) and therefore we cannot confirm the hypothesis $\mathcal{H}4$.

5.4 Participants preferences

To test $\mathcal{H}5$, we analyzed NASA-TLX scores, a direct preference question, and qualitative feedback.

NASA-TLX results (Fig. 11) showed that participants with normal vision favored LP across all dimensions. In contrast, participants with low vision reported higher physical and mental demand, together with more frustration for LP. Further analysis revealed this

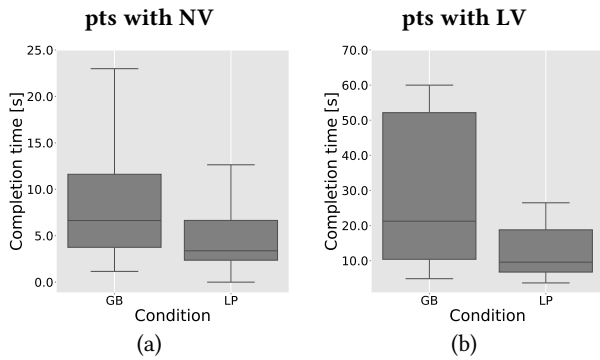


Figure 10: Completion time per condition (Task 2), only successful cases. Boxplots showing participants’ distributions, with (a) normal vision and (b) low vision.

was driven by participants with the most severe visual impairments (i.e., higher CPS). This finding requires careful contextualization: while producing LP based on CPS is ecologically valid, our strict requirement to render full headlines for experimental fairness between conditions imposed a magnification ceiling that frustrated these participants. In a real-world deployment, this could be mitigated by summarizing or truncating content to allow even higher magnification levels. However, for extreme visual constraints requiring extreme magnification levels, preserving a spatial 2D layout will inevitably become unfeasible, necessitating a shift to alternative solutions such as linear presentations (feed-like formats), as explored in prior work [5].

We triangulated these findings with objective performance (the sum of Task 1 and 2 completion times) and subjective preference. Based on total time, all participants (100%) performed better with LP. The final preference question aligned with NASA-TLX results for most participants (75%), though some discrepancies arose (25%). These discrepancies are not surprising, as NASA-TLX is a non-comparative tool known to be sensitive to task order and subjective recall. This subjectivity is likely amplified in participants with low vision, for whom the tasks were generally much more demanding, leading to less consistent ratings across conditions. To address this, we asked participants with low vision at the end of the study to indicate, for each NASA-TLX dimension, which condition they preferred. This post-hoc comparative measure captured their subjective preference, giving in fact a preference for LP for all NASA-TLX dimensions except effort and frustration, where there is no preference for either condition. This reinforces the overall advantage of LP. Nevertheless, a dedicated comparative workload assessment would provide a more reliable metric in future studies.

Finally, qualitative feedback provided additional insights:

- (1) **Comfort and Focus:** Participants favored LP for a more comfortable and focused experience. Several participants with low vision even changed their preference from GB to LP over the course of the experiment.
- (2) **Layout constraints:** Even if they overall prefer LP, they do not like to read headlines with too many lines and/or word breaks, which is something the relayouting method [21]

minimizes. This comment reinforces our idea of doing a re-layouting over the original newspaper page while increasing font size and not just the latter, but also opens opportunities to improve this relayouting algorithm.

- (3) **Desire for control:** A small minority (3 of 24, or 12.5%, which corresponds to the participants with the highest CPS) preferred GB despite their lower performance. This finding highlights that a subset of users may prioritize the sense of control provided by manual interaction, even when it imposes a significantly higher behavioral and performance cost by disrupting the structural perception of the layout.
- (4) **Feed-like layouts:** Participants enjoyed the vertical, feed-like layouts possible in LP. However, as argued in the Introduction, our goal is to support the complex, aesthetic layouts of realistic newspapers, for which simple feed-like designs are not a scalable solution, especially when ads or images are present, since they cannot be shrunk.

Taken together, these quantitative and qualitative results confirm our preference hypothesis $\mathcal{H}5$.

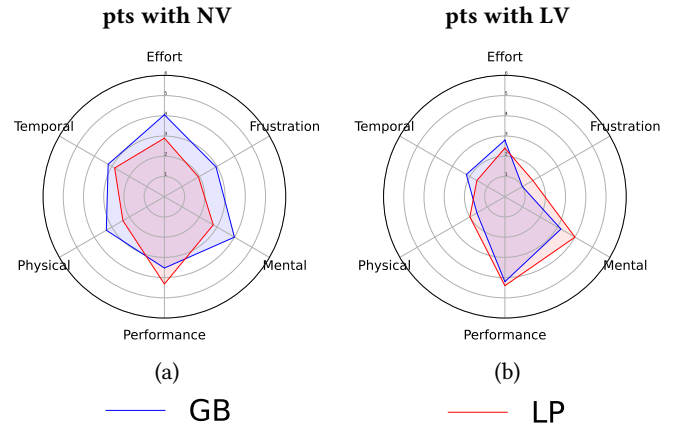


Figure 11: NASA-TLX average scores for each question, for both GB and LP, collected after completing the experiment. For all dimensions, lower scores indicate better perceived workload, except for performance, where higher scores are better. (a) Participants with normal vision show a clear advantage for LP. (b) Participants with low vision show similar trends, though results should be interpreted cautiously due to the small sample size, which can skew the results, and higher subjectivity of their responses, as tasks were generally more demanding and ratings less consistent across conditions (see text for more details).

5.5 Summary

Based on our previous results, let us make a summary focused in each hypothesis:

- $\mathcal{H}1$ Supported. LP significantly improved mechanical efficiency, with faster completion times and more efficient transitions between articles compared to the navigational overhead of GB.

H2 Supported. We established a reference path based on B trials, which followed a left-to-right, top-to-bottom reading order. We then measured how similar the reading paths in GB and LP were to this reference path. In both CVAS, LP paths were significantly more similar to the reference path than GB paths. This confirms that LP preserves the behavioral integrity of the reading process, whereas manual interaction forces a shift in the reader's exploration strategy.

H3 Supported. Locating a specific target headline was faster using LP, demonstrating the advantage of direct structural access in eliminating the navigational overhead and cost of fragmented manual interaction.

H4 Not supported. While we observed a non-significant trend related to target position, our analysis did not find sufficient statistical evidence to support this hypothesis.

H5 Supported. NASA-TLX scores, a direct preference question, and qualitative feedback all converged to show a strong overall user preference for LP.

5.6 Discussion

5.6.1 ConditionOrder impact. As shown in Table 2, we model conditions using a within-subject factor, which we counterbalance by systematic alternation to balance order (illustrated in Fig. 4). In addition, statistical analysis on ConditionOrder showed no significant results, indicating a lack of fatigue effects.

5.6.2 Generalization of the study. Our findings are expected to generalize within structured, layout-based documents, as loss of global context and interface manipulability are content-independent properties of GB vs. LP. While this study quantified text-only reading, the methodology could potentially be extended to richer layouts with additional entry points [42]. This would require identifying the specific natural reading paths associated with such layouts, which would differ from the one defined here. Furthermore, we recognise that future protocols would likely need adaptations to address this use case, such as moving beyond reading aloud or accounting for CPS limitations when integrating visual elements.

5.6.3 Training and familiarity effects. The potential role of training and familiarity also merits consideration; observed patterns were similar for participants with normal-vision (already familiar with magnification) and low vision, though effects were larger for the latter. This suggests that training could reduce difference magnitude for them, but their visual acuity will prevent eliminating the gap, and the qualitative pattern between GB and LP is unlikely to change.

6 Conclusion

In this work, we quantified the behavioral and performance cost of consuming digital layout-based media when visual access is constrained. This challenge affects both readers on small-screen devices and people with low vision.

Our multidimensional study provides empirical evidence and quantifies the impact of different newspaper interaction modalities. We found that while manual magnification (original edition with gesture-based magnification) is the standard, it imposes a navigational overhead that disrupts the reader's canonical strategy.

In contrast, large-print edition with direct access acts as a modality that restores behavioral integrity. Compared to gesture-based magnification, large-print edition with direct access enabled faster performance and a natural reading path (closely resembling the reference strategy), a stronger mental representation of document structure, and a significantly lower perceived workload.

These results indicate that magnification alone may be insufficient to support effective reading and consistent quality of experience in layout-based documents. The standard paradigm (original edition with gesture-based magnification) can make navigating a page more challenging, not due to a lack of features, but by potentially disrupting spatial context and forcing a shift from natural content consumption. By contrast, re-laying documents into large-print editions (large-print edition with direct access) preserves the spatial qualities of the layout (such as entry points) and supports efficient, predictable reading strategies.

Our findings offer important implications: designing digital media consumption tools should go beyond interactive magnification tools toward adaptive layouts that combine font scaling with structural re-flow. Our findings confirm that the effort required to generate large-print edition with direct access is justified not just by the improvements in reader performance, but also by the measurable restoration of behavioral integrity and satisfaction across diverse populations and devices, supporting efficient and inclusive access to newspapers, magazines, and other visually complex digital content. While our study focused on text-only layout-based documents, future work should extend these methods to richer layouts, incorporating graphical elements such as images and advertisements, and should include careful empirical studies to evaluate additional types of entry points.

Acknowledgments

This work was supported by the Association Nationale de la Recherche et de la Technologie (ANRT) through the CIFRE grant 2022/0927 in collaboration with Demain un Autre Jour.

References

- [1] Carlos Aguilar and Eric Castet. 2017. Evaluation of a gaze-controlled vision enhancement system for reading in visually impaired people. *PLOS ONE* 12, 4 (04 2017), 1–24. doi:10.1371/journal.pone.0174910
- [2] Mistral AI. 2025. MistralAI. <https://mistral.ai/>. Accessed: September 6th, 2025.
- [3] Ahmad Aljoudi. 2020. *Direct Manipulation of AI Text via Gestures*. Bachelor's Thesis. Malmö University, Faculty of Culture and Society (KS), School of Arts and Communication (K3).
- [4] Fernando Almeida and J. Monteiro. 2017. The role of responsive design in web development. *Webology* 14 (12 2017), 48–65.
- [5] Zainab Almeraj and Asmaa Alsumait. 2019. A User Centered Design Roadmap for Researchers and Designers Working with Visually Impaired and Blind Children. In *ACHI 2019: The Twelfth International Conference on Advances in Computer-Human Interactions*. IARIA, Athens, Greece.
- [6] Raheemat Atata. 2025. Evaluating Visual Hierarchies and Navigation Patterns to Improve Accessibility. *International Journal of Research Publication and Reviews* 6, 8 (Aug. 2025), 617–641. doi:10.55248/gengpi.6.0825.3080
- [7] Nilsu Atilgan, Ying-Zi Xiong, and Gordon E. Legge. 2020. Reconciling print-size and display-size constraints on reading. *Proceedings of the National Academy of Sciences* 117, 48 (2020), 30276–30284. doi:10.1073/pnas.2007514117
- [8] Max Bain, Jaesung Huh, Tengda Han, and Andrew Zisserman. 2023. WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. arXiv:2303.00747 [cs.SD] <https://arxiv.org/abs/2303.00747>
- [9] Paul J. Beckmann and Gordon E. Legge. 1996. Psychophysics of Reading—XIV. The Page Navigation Problem in Using Magnifiers. *Vision Research* 36, 22 (Nov. 1996), 3723–3733. doi:10.1016/0042-6989(96)00084-3
- [10] Alex Bowers, Allen Cheong, and Jan Lovie-Kitchin. 2007. Reading With Optical Magnifiers: Page Navigation Strategies and Difficulties. *Optometry and vision*

- science : official publication of the American Academy of Optometry 84 (02 2007), 9–20. doi:10.1097/01.opx.0000254035.39055.05
- [11] Raluca Budiu. 2015. Mobile User Experience: Limitations and Strengths. <https://www.nngroup.com/articles/mobile-ux/>. Accessed: August 25th, 2025.
 - [12] Mila Bujčić, Mikko Salminen, and Juho Hamari. 2023. More Immersed but Less Present: Unpacking Factors of Presence Across Devices. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences* (Nantes, France) (IMX '23). Association for Computing Machinery, New York, NY, USA, 140–149. doi:10.1145/3573381.3596152
 - [13] Thorsten Büring. 2001. *Zoomable User Interfaces on Small Screens - Presentation Interaction Design for Pen-Operated Mobile Devices*. PhD Thesis. University of Konstanz.
 - [14] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 981–992. doi:10.1145/2858036.2858498
 - [15] Aurélie Calabrèse, Allen M. Y. Cheong, Sing-Hang Cheung, Yingchen He, MiYoung Kwon, J. Stephen Mansfield, Ahalya Subramanian, Deyue Yu, and Gordon E. Legge. 2016. Baseline MNREAD Measures for Normally Sighted Subjects From Childhood to Old Age. *Investigative Ophthalmology & Visual Science* 57, 8 (July 2016), 3836. doi:10.1167/iovs.16-19580
 - [16] Andrew P. Chesham. 2003. *Poetry and eBooks: Managing Fixed Layouts in Reflowable Text*. Master's Thesis. University of Simon Fraser.
 - [17] Paul T. Chiou, Robert Winn, Ali S. Alotaibi, and William G. J. Halfond. 2024. Automatically Detecting Reflow Accessibility Issues in Responsive Web Pages. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*. ACM, New York, NY, USA, 1–13. doi:10.1145/3597503.3639229
 - [18] Michael Christen and Mathias Abegg. 2017. The Effect of Magnification and Contrast on Reading Performance in Different Types of Simulated Low Vision. *Journal of Eye Movement Research* 10 (05 2017). doi:10.16910/jemr.10.2.5
 - [19] Shao Jing Ding, Ernest Tak Hei Lam, Dickson KW Chiu, Mavis Man wai Lung, and Kevin KW Ho. 2021. Changes in reading behaviour of periodicals on mobile devices: A comparative study. *Journal of Librarianship and Information Science* 53, 2 (2021), 233–244. arXiv:https://doi.org/10.1177/0961000620938119 doi:10.1177/0961000620938119
 - [20] Sukru Eraslan, Yeliz Yesilada, and Simon Harper. 2015. Eye tracking scanpath analysis techniques on web pages: A survey, evaluation and comparison. *Journal of Eye Movement Research* 9, 1 (2015), 2. doi:10.16910/jemr.9.1.2
 - [21] Sebastian Gallardo, Maria Cristina Riff, Dorian Mazauric, and Pierre Kornprobst. 2023. Newspaper Magnification with Preserved Entry Points. (Sept. 2023). <https://hal.science/hal-04210840> working paper or preprint.
 - [22] Christina Granquist, Yueh-Hsun Wu, Rachel Gage, Michael D. Crossland, and Gordon E. Legge. 2018. How People with Low Vision Achieve Magnification in Digital Reading. *Optometry and Vision Science* 95, 9 (Sept. 2018), 711–719. doi:10.1097/OPX.0000000000001261
 - [23] André Hajek, Razak Gyasi, Benedikt Kretzler, and Hans-Helmut König. 2023. Vision and hearing problems and psychosocial outcomes: longitudinal evidence from the German Ageing Survey. *Social Psychiatry and Psychiatric Epidemiology* 59 (11 2023), 1–8. doi:10.1007/s00127-023-02588-9
 - [24] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. doi:10.1177/154193120605000909
 - [25] Seongsil Heo, Roberto Manduchi, and Suzana Chung. 2024. Reading with Screen Magnification: Eye Movement Analysis Using Compensated Gaze Tracks. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*. ACM, Glasgow United Kingdom, 1–6. doi:10.1145/3649902.3656493
 - [26] Barry A. Hollander, Dean M. Krugman, Tom Reichert, and J. Adam Avant. 2011. The E-Reader as Replacement for the Print Newspaper. *Pub Res Q* 27 (2011), 126–134. doi:10.1007/s12109-011-9205-8
 - [27] Nils Holmberg. 2004. *Eye movement patterns and newspaper design factors: An experimental approach*. Master's Thesis. Lund University, Cognitive Science, Lund, Sweden.
 - [28] Kenneth Holmqvist and Constanze Wartenberg. 2005. *The role of local design factors for newspaper reading behaviour: An eye-tracking perspective*. LUCS Report 127. Lund University, Cognitive Studies, Lund, Sweden. ISSN 1101-8453.
 - [29] Jana Holsanova, Henrik Rahm, and Kenneth Holmqvist. 2006. Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements. *Visual Communication* 5, 1 (Feb. 2006), 65–93. doi:10.1177/1470357206061005
 - [30] Kasper Hornbæk, Benjamin B. Bederson, and Catherine Plaisant. 2002. Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Trans. Comput.-Hum. Interact.* 9, 4 (Dec. 2002), 362–389. doi:10.1145/586081.586086
 - [31] Robert J.K. Jacob and Keith S. Karn. 2003. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The Mind's Eye*, J. Hyönä, R. Radach, and H. Deubel (Eds.). North-Holland, Amsterdam, 573–605. doi:10.1016/B978-044451020-4/50031-1
 - [32] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction* (second edition ed.). Morgan Kaufmann Publishers, an imprint of Elsevier, Cambridge, Massachusetts.
 - [33] Gordon Legge. 2016. Reading Digital with Low Vision. *Visible language* 50 (08 2016), 102–125.
 - [34] Gordon E. Legge. 2006. *Psychophysics of reading in normal and low vision / Gordon E. Legge and colleagues*. Lawrence Erlbaum, Mahwah, NJ.
 - [35] Gordon E. Legge, John A. Ross, L. Michael Isenberg, and J. M. LaMay. 1992. Psychophysics of Reading—Clinical Predictors of Low-Vision Reading Speed. *Investigative Ophthalmology & Visual Science* 33, 3 (1992), 677–687.
 - [36] J. Stephen Mansfield, Sonia J. Ahn, Gordon E. Legge, and Andrew Luebker. 1993. Commentary on Section 4 - A New Reading-Acuity Chart for Normal and Low Vision, In Noninvasive Assessment of the Visual System. *Noninvasive Assessment of the Visual System*, NSuD.3. doi:10.1364/NAVS.1993.NSuD.3
 - [37] Ethan Marcotte. 2010. Responsive Web Design. <https://alistapart.com/article/responsive-web-design/>. Accessed: December 3rd, 2025.
 - [38] Kate Moran. 2016. Reading Content on Mobile Devices. <https://www.nngroup.com/articles/mobile-content/>. Accessed: August 18th, 2025.
 - [39] Jakob Nielsen. 2011. Mobile Content Is Twice as Difficult. <https://www.nngroup.com/articles/mobile-content-is-twice-as-difficult-2011/>. Accessed: August 25th, 2025.
 - [40] Denis Nikitenko, Jordan Evans, David R. Flatla, Thomas Driscoll, Graham Quinlan, and Kyle Lukaszek. 2024. Situational visual impairments on mobile devices - modeling the effects of bright outdoor environments. In *Proceedings of the 50th Graphics Interface Conference* (Halifax, NS, Canada) (GI '24). Association for Computing Machinery, New York, NY, USA, Article 1, 10 pages. doi:10.1145/3670947.3670959
 - [41] Kenton O'Hara and Abigail Sellen. 1997. A comparison of reading paper and on-line documents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) (CHI '97). Association for Computing Machinery, New York, NY, USA, 335–342. doi:10.1145/258549.258787
 - [42] Đurdana Ozretić Došen and Lidija Brkljačić. 2018. Key design elements of daily newspapers: Impact on the reader's perception and visual impression. *KOME* (Aug. 2018). doi:10.17646/KOME.75692.93
 - [43] Ar Poorva Priyadarshini. 2024. The Impact of User Interface Design on User Engagement. *International Journal of Engineering Research* 13, 03 (2024).
 - [44] R Project. 2025. R: The R Project for Statistical Computing. <https://www.r-project.org/>. Accessed: September 6th, 2025.
 - [45] Zhanna Sarsenbayeva, Niels van Berkel, Chu Luo, Vassilis Kostakos, and Jorge Goncalves. 2017. Challenges of situational impairments during interaction with mobile devices. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (Brisbane, Queensland, Australia) (OzCHI '17). Association for Computing Machinery, New York, NY, USA, 477–481. doi:10.1145/3152771.3156161
 - [46] Somipam R. Shimray, Chennupati Keerti, and Chennupati K. Ramaiah. 2015. An Overview of Mobile Reading Habits. *DESIDOC Journal of Library & Information Technology* 35, 5 (2015), 343–354.
 - [47] Shneiderman. 1983. Direct Manipulation: A Step Beyond Programming Languages. *Computer* 16, 8 (Aug. 1983), 57–69. doi:10.1109/MC.1983.1654471
 - [48] Meini Tang, Roberto Manduchi, Susana Chung, and Raquel Prado. 2023. Screen Magnification for Readers with Low Vision: A Study on Usability and Performance. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York NY USA, 1–15. doi:10.1145/3597638.3608383
 - [49] Garreth W. Tigwell, Rachel Menzies, and David R. Flatla. 2018. Designing for Situational Visual Impairments: Supporting Early-Career Designers of Mobile Content. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) (DIS '18). Association for Computing Machinery, New York, NY, USA, 387–399. doi:10.1145/3196709.3196760
 - [50] The New York Times. 1967. Times Begins an Edition Printed in Large Type. <https://www.nytimes.com/1967/03/06/archives/times-begins-an-edition-printed-in-large-type.html>. Accessed: August 14th, 2025.
 - [51] M. Valsecchi, K. R. Gegenfurtner, and A. C. Schutz. 2013. Saccadic and smooth-pursuit eye movements during reading of drifting texts. *Journal of Vision* 13, 08 (Aug. 2013), 8–8. doi:10.1167/13.10.8
 - [52] Radu-Daniel Vatavu. 2021. Accessibility of Interactive Television and Media Experiences: Users with Disabilities Have Been Little Voiced at IMX and TVX. In *Proceedings of the 2021 ACM International Conference on Interactive Media Experiences* (Virtual Event, USA) (IMX '21). Association for Computing Machinery, New York, NY, USA, 218–222. doi:10.1145/3452918.3465485
 - [53] W3C Web Accessibility Initiative (WAI). 2018. Understanding SC 1.4.10: Reflow (Level AA). <https://www.w3.org/WAI/WCAG21/Understanding/reflow.html>. World Wide Web Consortium (W3C). Accessed 2025-08-13. Understanding document for WCAG 2.1, SC 1.4.10.
 - [54] Peng Wang, Dickson K.W. Chiu, Kevin K.W. Ho, and Patrick Lo. 2016. Why read it on your mobile device? Change in reading habit of electronic magazines for university students. *The Journal of Academic Librarianship* 42, 6 (2016), 664–669. doi:10.1016/j.acalib.2016.08.007

- [55] Ru Wang, Zach Potter, Yun Ho, Daniel Killough, Linxiu Zeng, Sanbrita Mondal, and Yuhang Zhao. 2024. GazePrompt: Enhancing Low Vision People's Reading Experience with Gaze-Aware Augmentations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–17. doi:10.1145/3613904.3642878
- [56] Ru Wang, Linxiu Zeng, Xinyong Zhang, Sanbrita Mondal, and Yuhang Zhao. 2023. Understanding How Low Vision People Read Using Eye Tracking. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. doi:10.1145/3544548.3581213
- [57] Ying-Zi Xiong, Nilsu Atilgan, Donald C. Fletcher, and Gordon E. Legge. 2022. Digital Reading with Low Vision: Principles for Selecting Display Size. *Optometry and Vision Science* 99, 8 (Aug. 2022), 655–661. doi:10.1097/OPX.0000000000001919
- [58] Kun Yue, Mingshan Zhang, Jingruo Chen, Chun Yu, Kexin Nie, Zhiqi Gao, Jinghan Yang, Chen Liang, and Yuanchun Shi. 2024. SituFont: A Just-in-Time Adaptive Intervention System for Enhancing Mobile Readability in Situational Visual Impairments. (12 Oct. 2024). arXiv:2410.09562 <https://arxiv.org/abs/2410.09562>
- [59] Li Yujian and Liu Bo. 2007. A Normalized Levenshtein Distance Metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 6 (2007), 1091–1095. doi:10.1109/TPAMI.2007.1078
- [60] Daniela Zambarbieri, Elena Carniglia, and Carlo Robino. 2008. Eye Tracking Analysis in Reading Online Newspapers. *Journal of Eye Movement Research* 2, 4 (Nov. 2008), 7. doi:10.16910/jemr.2.4.7

Appendix: Linear-mixed effects model ANOVA tables

<i>Task 1 success ratio - Fixed effects (type III tests)</i>					
Effect	Num DF	Den DF	Statistic	p-value	Effect Size
NewspaperID	5	215.94	$F = 5.72$	<.001	$\eta_p^2 = 0.12$
CVAS (Normal vs Low)	1	23.71	$F = 20.92$	<.001	$\eta_p^2 = 0.55$
Content Version (1,2,3)	2	223.45	$F = 3.72$	0.025	$\eta_p^2 = 0.03$
NewspaperID \times CVAS	5	216.1	$F = 3.83$.002	$\eta_p^2 = 0.08$

Table 3: Task 1 success ratio

<i>Task 1 completion time - Fixed effects (type III tests)</i>					
Effect	Num DF	Den DF	Statistic	p-value	Effect Size
Condition (GB vs LP)	1	209.43	$F = 116.5$	<.001	$\eta_p^2 = 0.42$
NewspaperID	5	209.89	$F = 9.57$	<.001	$\eta_p^2 = 0.19$
CVAS (Normal vs Low)	1	18.61	$F = 50.19$	<.001	$\eta_p^2 = 0.73$
Content Version (1,2,3)	2	213.5	$F = 3.21$	0.04	$\eta_p^2 = 0.03$
Condition \times CVAS	1	209.18	$F = 146.3$	<.001	$\eta_p^2 = 0.41$

Table 4: Task 1 completion time

<i>Task 1 reading path similarity - Fixed effects (type III tests)</i>					
Effect	Num DF	Den DF	Statistic	p-value	Effect Size
Condition (GB vs LP)	1	210.75	$F = 11.98$	<.001	$\eta_p^2 = 0.14$
NewspaperID	5	212.51	$F = 6.15$	<.001	$\eta_p^2 = 0.13$
Reading frequency	3	212.49	$F = 5.94$	<.001	$\eta_p^2 = 0.08$
<i>Post-hoc comparison</i>					
NewspaperID (Layout #19 vs others)	5	209.19	$F = 8.93$	<.001	$\eta_p^2 = 0.18$

Table 5: Task 1 reading path similarity

<i>Task 1 reading time - Fixed effects (type III tests)</i>					
Effect	Num DF	Den DF	Statistic	p-value	Effect Size
Condition (GB vs LP)	1	209.24	$F = 6.02$	0.01	$\eta_p^2 = 0.05$
CVAS (Normal vs Low)	1	18.3	$F = 15.5$	<.001	$\eta_p^2 = 0.46$
Trial Index (1...6)	1	209.13	$F = 7.66$.006	$\eta_p^2 = 0.04$
Content Version (1,2,3)	2	211.57	$F = 3.72$	0.02	$\eta_p^2 = 0.03$
Condition \times NewspaperID	5	209.02	$F = 4.97$	<.001	$\eta_p^2 = 0.11$

<i>Task 1 transition time - Fixed effects (type III tests)</i>					
Effect	Num DF	Den DF	Statistic	p-value	Effect Size
Condition (GB vs LP)	1	210.09	$F = 73.57$	<.001	$\eta_p^2 = 0.29$
CVAS (Normal vs Low)	1	19.58	$F = 71.6$	<.001	$\eta_p^2 = 0.8$
NewspaperID	5	211.23	$F = 7.45$	<.001	$\eta_p^2 = 0.15$
Trial Index (1...6)	1	209.62	$F = 4.38$.037	$\eta_p^2 = 0.02$
Condition \times CVAS	1	209.47	$F = 113.82$	<.001	$\eta_p^2 = 0.35$

Table 6: Task 1 reading and transition time

<i>Task 2 success ratio - Fixed effects (type III tests)</i>					
Effect	Num DF	Den DF	Statistic	p-value	Effect Size
NewspaperID	5	219.47	$F = 4.10$.001	$\eta_p^2 = 0.09$
NewspaperID \times CVAS	5	219.66	$F = 3.30$.006	$\eta_p^2 = 0.07$

Table 7: Task 2 success ratio

<i>Task 2 completion time - Fixed effects (type III tests)</i>					
Effect	Num DF	Den DF	Statistic	p-value	Effect Size
Condition	1	212.63	$F = 15.06$	<.001	$\eta_p^2 = 0.12$
CVAS (Normal vs Low)	1	23.55	$F = 47.20$	<.001	$\eta_p^2 = 0.71$
Condition \times CVAS	1	210.63	$F = 14.81$	<.001	$\eta_p^2 = 0.07$

Table 8: Task 2 completion time