



Text readability in augmented reality: a multivocal literature review

Maxime Cauz¹ · Antoine Clarinval¹ · Bruno Dumas¹

Received: 7 September 2023 / Accepted: 15 January 2024 / Published online: 4 March 2024
© The Author(s) 2024

Abstract

Augmented reality (AR) is making its way into many sectors. Its rapid evolution in recent years has led to the development of prototypes demonstrating its effectiveness. However, to be able to push these prototypes to the scale of fully usable applications, it is important to ensure the readability of the texts they include. To this end, we conducted a multivocal literature review (MLR) to determine the text parameters a designer can tune, as well as the contextual constraints they need to pay attention to, in relation to Optical See-Through (OST) and Video See-Through (VST) displays. We also included guidelines from device manufacturing and game engines sites to compare the current state of research in the academic and industrial worlds. The results show that parameters pertaining more to letter legibility have been extensively studied (e.g., color and size), while those pertaining to the whole text still require further research (e.g., alignment or space between lines). The former group of parameters, and their associated constraints, were assembled in the form of two decision trees to facilitate implementation of AR applications. Finally, we also concluded that there was a lack of alignment between academic and industrial recommendations.

Keywords Text · Readability · Legibility · Augmented Reality · Mixed Reality

1 Introduction

Augmented Reality (AR) is a burgeoning field that combines virtual elements with the real world, registered in three dimensions and interactive in real time (Azuma 1997). Led by the latest technological advances, AR has gained significant attention in recent years. It encompasses a broad range of applications, spanning across industries such as gaming, education, healthcare, architecture, and manufacturing (Cárdenas-Robledo et al. 2022; Liberatore and Wagner 2021). Head-Mounted Devices (HMDs) play a crucial role in enabling augmented reality experiences. These devices, in the form of glasses or helmets, are designed to be worn on the user's head, providing a wearable interface for AR. Nowadays, two technologies exist in terms of HMDs, each with

advantages and disadvantages (Schmalstieg and Höllerer 2016). On the one hand, the Optical See-Through (OST) display generates the virtual elements in the form of light projected on a transparent screen, where it is mixed with the light coming from the user's surroundings before being perceived by the user's eyes. OST displays do not degrade the reality but are more sensible to latency since the real and virtual information would no longer be correctly aligned. In addition, ambient luminance strongly impacts OST displays (Gattullo et al. 2015). The reality modifies the final rendering of the virtual world due to the combination with the real world. Thus, the perceptible spectrum of virtual colors will depend on the background illuminance. The brighter the environment, the higher the intensity of the color of the virtual world should be to be distinguishable. In the same way, dark colors are transparent on OST displays. On the other hand, the Video See-Through (VST) display captures the reality with a camera and mixes it with the virtual elements by video computation before rendering the result on a screen in front of the user's eyes. VST displays, which relies on both screen resolution and the camera's real-world capture quality, are more susceptible to inducing motion sickness compared to OST displays (Munafò et al. 2016; Yuan et al. 2018; Kaufeld et al. 2022). These new approaches to

✉ Maxime Cauz
maxime.cauz@unamur.be

Antoine Clarinval
antoine.clarinval@unamur.be

Bruno Dumas
bruno.dumas@unamur.be

¹ Namur Digital Institute, University of Namur, Namur, Belgium

displaying information raise the question of whether User Interface (UI) guidelines established for traditional computer systems also apply to these types of devices. A UI is composed of various visual elements such as panels, buttons, images, text, etc. For text in particular, Kim et al. (2014) demonstrated that HMDs have their own readability characteristics, making it necessary to establish new guidelines.

To address this issue, the first thing to consider is the presentation of the assets that make up AR interfaces. To this end, numerous studies have investigated various parameters in controlled experiments. In this paper, we focus on text, which presents numerous parameters (e.g., font, size, color, position), making it challenging for designers. In addition, text is one of the most widely used assets in AR applications. It can be used to label objects or places, to give instructions or details about real and virtual elements. In a study devoted to maintenance, assembly, and training, Gattullo et al. (2022) determined that text represents 26% of the assets used in the 122 papers they reviewed. As far as we found, there exists only one recent survey on the topic. Erickson et al. (2020a) reviewed the literature to explore user's perception of text on OST HMDs. Their results summarized the literature they found on text drawing style, text color, background texture, environment lighting, text position, and text size. However, they restrained their search to the IEEEExplore and ACM DL digital libraries and to OST displays. In addition, Korkut and Surer (2023) recently reported a need to extend text-related guidelines to Virtual Reality, thus to VST displays.

We conducted a multivocal literature review that extends the work of Erickson et al. (2020a) by (1) opening the search beyond the IEEEExplore and ACM DL digital libraries, (2) including works on VST HMDs to facilitate the comparison between the two types of displays and expand the applicability of our results, and (3) including official documentations from manufacturers or game engines as grey literature. Ultimately, the aim of our research is to identify the text parameters designers can tune to improve text readability, as well as the associated contextual constraints, in order to propose an up-to-date state of knowledge and a designing guide grounded in both academic and grey literature. The guide is of interest both to researchers wishing to explore the readability of text, and to designers wishing to set up functional applications adapted to their users, their tasks, and their environment. Therefore, we formulated the following research questions concerning AR applications on HMDs:

RQ1: What are the different text parameters that can be tuned and how to tune them to maximize the readability of a text?

RQ2: What are the contextual constraints to consider when designing an application including texts?

RQ3: What is the gap between the academic and grey literature in terms of the guidelines proposed on text readability?

The remaining of this paper is organized as follows. In Sect. 2, we present the methodology guiding the conduct of the multivocal literature review, following the steps of Garousi et al. (2019). The need for a review and the research questions were already established in the Introduction. The next main steps consist of developing the protocol (Sect. 2.1, 2.2, 2.3, and 2.4), conducting the review (Sect. 2.6), and assessing the quality of the coverage (Sect. 2.5). In Sect. 3, we report the results of the selected papers for each text parameter we identified. In Sect. 4, building on these results, we develop our designing guide in the form of decision trees built from those parameters. Finally, in Sect. 5, we discuss the implications of our results, limitations, and future works before concluding in Sect. 6.

2 Method

We conducted the Multivocal Literature Review (MLR) by following the guidelines from Garousi et al. (2019). They adapted the work of Kitchenham et al. (2015), which proposed guidelines to conduct Systematic Literature Review (SLR) in software engineering, to present their MLR protocol that includes Grey Literature. Overall, the protocol remains the same, except that in addition to targeting academic libraries (e.g., IEEEExplore, ACM DL digital libraries), those containing Grey Literature (e.g., Google) are also considered.

2.1 Grey literature sources

Garousi et al. (2019) noted that the most widely accepted definition of the Grey Literature is the Luxembourg definition of Schopfel and Farace (2010): “< Grey Literature > is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers, i.e., where publishing is not the primary activity of the producing body”. According to them, integrating Grey Literature in a survey allows to “close the gap between the academic research and professional practice”, which is needed to answer RQ3, and benefits from the knowledge of both academia and practice, which we need to answer RQ1 and RQ2. In addition, Lawrence et al. (2014) argued that a part of the Grey Literature equally follows a peer-review process, ensuring its quality. To this end, Adams et al. (2017) divided the Grey Literature in three tiers based on the source expertise and outlet control dimensions. The first tier consists of books, magazines, or even reports with a significant control and credibility. The second tier consists of question/answer forums, wikis, videos, etc. with a moderate control and credibility. Finally, the last tier consists of blogs, tweets, letters, etc. with low or no control and credibility.

In order to ensure a certain quality of documents, the search for Grey Literature targets device manufacturers or game engine websites. They correspond to the first tier of Grey Literature defined by Adams et al. (2017) (i.e., high control and credibility). Indeed, manufacturers inevitably asked themselves the question of text legibility when faced with the task of designing their device's operating system and applications. In addition, we assume that manufacturers and game engine teams whether conducted experiments to validate their documentation before publishing it or relied on scientific literature. Then, we decided to search for guidelines on the official documentation of Microsoft, Google Fonts (named Google in the rest of the paper), Valve (HTC Vive), Meta (Oculus), Lynx, Varjo, Magic Leap, Unity, and Unreal Engine. The first seven are major device manufacturers, while the last two are the two main game engines often used to develop AR/VR applications. We conducted the search manually by screening the hierarchy of their documentation for developers. Note that manufacturers focus on the documentation of their devices. In the same way, game engines focus on the documentation of devices they support.

2.2 Digital libraries and search terms

According to the online Cambridge Dictionary, legibility is defined as “the fact of being *easy to read*, or the degree to which something is easy to read”.¹ In contrast, readability is defined as “the quality of being *easy and enjoyable to read*”.² In other words, readability extends the more restricted legibility concept with User Experience (UX) considerations. Nevertheless, we observed that the two terms are sometimes interchanged in the literature. In addition, the definition of readability is sometimes extended to include the ability to understand text as with language research. In this survey, we rely on the definition of text readability as given by the Cambridge Dictionary, but we included legibility in our search as well since they are used interchangeably in the literature.

Regarding White Literature, we focus our initial search on the IEEEExplore, ACM DL, and ScienceDirect digital libraries. All three are popular among researchers related to computer science domains (Kitchenham et al. 2015). In addition, the papers selected in the work of Erickson et al. (2020a) were also included. Three concepts are at the center of our research: the asset, the studied property, and the application domain. Except for the last concept, we decided to restrain our search terms to as few as possible to prevent digital libraries from returning too many non-related papers. As

discussed by Erickson et al. (2020a), the terms related to the asset and studied property are common to various domains, which rapidly increases the number of articles yielded by an automated search. For example, the term “visual” can take on different meanings depending on the terms with which it is associated. However, we mitigated this limitation by complementing the search query with other literature retrieving techniques (i.e., snowballing and coverage assessment). The search terms were applied to the title, abstract, and keywords of papers in the literature. All articles published before the end of 2022 have been considered. The complete search query is given below:

Text AND (Readability OR Legibility) AND (“Augmented Reality” OR AR OR “Mixed Reality” OR MR OR “Virtual Reality” OR VR OR “Head-mounted display” OR HMD OR “Head-worn display” OR HWD).

2.3 Inclusion and exclusion criteria

To assess the relevance of the identified sources, we defined inclusion and exclusion criteria. A document is relevant if it satisfies the inclusion criteria and none of the exclusion criteria. The only inclusion criterion is defined as follows: the document must focus on text readability in AR or VR on HMDs. The exclusion criteria and their rationale are as follows:

1. *Not written in English*: The standard language for white literature is English.
2. *Duplicated*: We consider two documents as duplicate only if they present the same research published at the same venue (i.e., our search may return the same article multiple times due to cross-references in digital libraries) or if the authors clearly indicate that two papers are duplicate.
3. *[For White Literature] Not peer-reviewed*: We consider only documents published in conferences or journal outlets that mandate peer-reviewing.
4. *[For White Literature] No evaluation of the text parameters*: By requiring documents to include an evaluation, we ensure that they present novel research based on findings validated with users.
5. *[For Grey Literature] No official documentation of an HMD manufacturer or a game engine used to develop AR/VR applications*: As discussed in Sect. 2.1, we restrain our search to sources with high control and credibility.

2.4 Snowballing phase

We conducted a phase of reverse and forward snowballing on the relevant papers yielded by the search query. For the forward method, we obtained the citations on Google

¹ <https://dictionary.cambridge.org/dictionary/english/legibility?q=Legibility> (Accessed 26 November 2023).

² <https://dictionary.cambridge.org/dictionary/english/readability?q=Readability> (Accessed 26 November 2023).

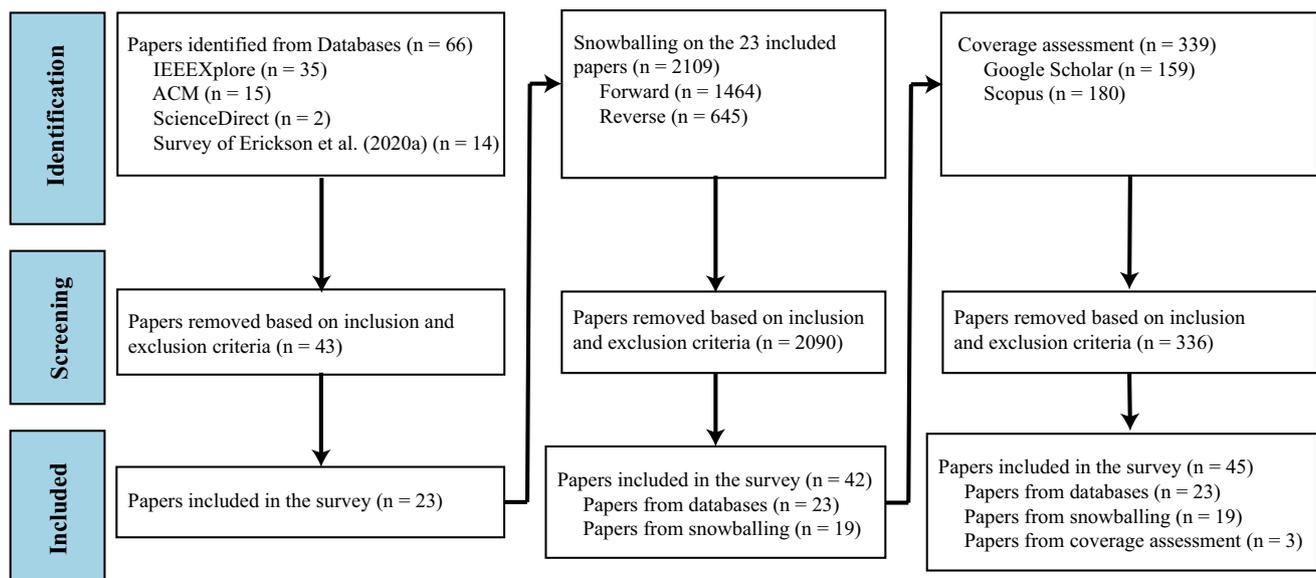


Fig. 1 Summary of the White Literature collection process. (Created with Adobe Illustrator)

Scholar and Scopus. We did not repeat the snowballing phase since the majority of relevant papers returned by this phase were already obtained at the beginning of the process. The longer the process went on, the fewer papers appeared that matched our criteria. Additionally, we ended the survey with a coverage check phase, which assesses the completeness of our protocol.

2.5 Coverage check

To assess the completeness of the review, we conducted a keyword search on three popular indexers, namely Google, Google Scholar, and Scopus. The first one assesses coverage for the Grey Literature, while the last two assess coverage for the White Literature. To this end, we updated our search query defined in Sect. 2.2 to include selected works that do not match it. For each group of terms associated with the same concept that a paper does not match in its title, abstract or keywords, we searched the terms it uses to discuss on the concept. Based on the terms obtained, we selected those which best present the concept and appear the most. However, we would have liked to add the words “reading” and “visual”, but for the same reason as explained in Sect. 2.2, they yield too many non-related works. Therefore, the search query used on Google indexer combined the name of each company with a synonym of the word “Text” or “Readability” (see terms used below) or “UI interface” or “Guidelines” or “Best practices”. The last three are common terms in developer documentations. The final search query used to conduct the coverage check on Google Scholar and Scopus indexers was:

(Text OR Typography OR Typeface OR Characters OR Fonts) AND (Readability OR Legibility) AND (“Augmented Reality” OR AR OR “Mixed Reality” OR MR OR “Virtual Reality” OR VR OR “Head-mounted display” OR HMD).

Nevertheless, considering the important number of results that Google Scholar usually returns, we set a limit of exploration to 100 consecutive papers excluded. For the search on Google, we limited to five pages.

2.6 Conducting the review

We began by collecting papers from the White Literature. The process is illustrated in Fig. 1. Based on the search query specified in Sect. 2.2, we obtained 52 results published before the end of 2022: 35 from IEEEExplore, 15 from ACM DL, and 2 from ScienceDirect. To these must be added the papers selected in the review by Erickson et al. (2020a), resulting in a total of 66. We applied the inclusion and exclusion criteria and kept 23 of these papers for our survey. Then, we conducted a phase of snowballing as explained in Sect. 2.4. We collected 2,109 papers in total, 645 papers from the reverse snowballing and 1,464 from the forward. We applied the inclusion and exclusion criteria and included 19 additional papers. Finally, we conducted an additional coverage check on two general academic digital libraries to capture potentially missed papers (see Sect. 2.5). This process returned 339 papers, 3 of which were added to the set. The final set included in the survey contains 45 papers.

Regarding Grey Literature, we found guidelines on text rendering only in the Microsoft (2023a,b,c), Google (2017, 2023), and Oculus (2023a; b) documentation. The additional search on the Google indexer returned no additional results

except websites that do not match our criterion of a device manufacturer or game engine.

3 Results by text parameter

This section summarizes the guidelines coming from the selected papers and pages of Grey Literature. We recorded 24 articles on OST, 17 on VST, and 4 on both types of screens. Table 1 lists the 45 papers by indicating their domain, their types of devices, and the parameters of the text studied. In this section, we classified them according to the text parameter they focus on. We carefully read each paper and each page of Grey Literature to determine the text parameters that were being studied. Once the list was obtained, we tagged each paper as shown in Table 1. We defined 7 parameters: the color, the text drawing style, the font-style, the font-size, the anchor, the position (including depth), and the text appearance and segmentation. The last parameter is a category that includes all parameters less frequently discussed in the literature such as text length, line space, and text weight. Finally, we read all the papers and pages of Grey Literature to extract atomic guidelines and grouped those that are the same. Nevertheless, although contrast is not a text parameter since depending on the color, it will be discussed in first to bring up some important key points for understanding the text drawing style and color parameters. The information of each section is presented in the following order. First, useful concepts and results shared between the two types of displays and coming from the White Literature are discussed. Second, results limited to one type of display are described. Last, the information coming from the Grey Literature is addressed.

3.1 Contrast

Since Augmented Reality (AR) consists of the combination of the real and virtual worlds, two of the environmental factors that affect readability are the background texture and the background illuminance. Leykin and Tuceryan (2004) demonstrated that background variations only affect readability when the text contrast is low. Debernardis et al. (2014) and Gattullo et al. (2014, 2015) observed the need for a minimal contrast ratio between the text and the background on the two display types. However, because of their intrinsic rendering characteristics, the two display types react differently to the contrast requirement between the text and the background. On VST displays, the background illuminance is normalized by the camera. Therefore, the background illuminance strongly affects OST displays but not VST displays (Gattullo et al. 2014, 2015). On OST displays, the brighter the background, the less the colors will appear, as discussed in Sect. 1.

Specific to VST displays – Kojic et al. (2020) recommended a minimal ratio of 7:1 as proposed by the Web Content Accessibility Guidelines 2.0 (WCAG2.0)³ and to avoid using a too high contrast ratio. They hypothesized that bright color on a screen as close to the eyes will emit too much light. The formula they used to calculate the contrast ratio was the one from the WCAG2.0,⁴ based on the relative luminance of the foreground and background colors (i.e., calculated on the quantity of red, green, and blue in the linear RGB color space).

Specific to OST displays – Blanc-Goldhammer and MacKenzie (2018) suggested a minimal ratio of 1.6:1 calculated as the ratio between the text luminance and the background luminance (i.e., the real luminance emitted by the two elements).

From Grey Literature – As expected, the three companies recommend a minimal contrast on the two types of display and note the impact of background illuminance on OST displays. They also suggest preventing too bright experiences.

3.2 Text drawing style

Through the included papers, four text drawing styles were evaluated: the billboard (see Fig. 2a), the outline (see Fig. 2b), the drop shadow (see Fig. 2c), and the plain text styles (see Fig. 2d). The billboard style employs a plain color box behind the text to isolate it from the influence of the background. Since the occlusion created by the box must not be avoided, it is recognized as the most effective style (Gabbard et al. 2006, 2007; Gattullo et al. 2015; Fiorentino et al. 2013). In comparison, the outline (resp. drop shadow) style draws a border (resp. shadow) around the letters of the text. Lastly, for the plain text styles on both display types, the efficiency depends on the chosen color, the contrast between the color and the background, and, for OST displays, the color illuminance (i.e., prefer a color with a high illuminance) (Gabbard et al. 2006, 2007; Gattullo et al. 2014, 2015).

Specific to VST displays – Kojic et al. (2022) observed the users' preference for the flat billboard (see Fig. 3a) but a better performance on the curved one (see Fig. 3b). Grout et al. (2015) identified a positive correlation between the interest in the curvature and the size of the text sample used. Wei et al. (2020) determined that the curve angle must be at maximum between 50° and 60° around one axis rather than two. Concerning the outline style, adding a minimal outline (e.g., 1 px) improves readability, but increasing it too much

³ <http://www.w3.org/TR/2008/REC-WCAG20-20081211/#visual-audio-contrast7> (Accessed 26 November 2023).

⁴ <https://www.w3.org/TR/WCAG20-TECHS/G17.html> (Accessed 26 November 2023).

Table 1 The environment, the devices, and the text parameters studied in each selected paper

	Augmented reality	Virtual reality	Optical see-through	Video see-through	Color and/or contrast	Text drawing style	Font type	Font size	Anchoring	Position	Text appearance and segmentation
Gabbard et al. (2006)	X		X		X	X				X	
Gabbard et al. (2007)	X		X		X	X					
Renkewitz et al. (2008)	X		X		X		X				
Gabbard et al. (2010)	X		X		X	X					
Fiorentino et al. (2013)	X		X		X	X					
Sridharan et al. (2013)	X		X		X	X					
Hincapie-Ramos (2014)	X		X		X	X					
Hincapie-Ramos (2015)	X		X		X	X					
Borg et al. (2015)	X		X		X	X	X	X			
Zhao et al. (2017)	X		X		X	X	X				
Rzayev et al. (2018)	X		X		X				X		X
Blanc-Goldhammer and MacKenzie (2018)	X		X		X			X			
Chang et al. (2019)	X		X							X	
Klose et al. (2019)	X		X					X			
Kim et al. (2019)	X		X		X					X	
Gabbard et al. (2019)	X		X		X					X	
Woodward et al. (2020)	X		X					X		X	
Fukushima et al. (2020)	X		X					X		X	
Rosilius et al. (2021)	X		X		X						
Erickson et al. (2021)	X		X		X					X	
Falk et al. (2021)	X		X		X	X					X
Koide et al. (2022)	X		X							X	
Arefin et al. (2022)	X		X							X	
Lee et al. (2023)	X		X						X		
Gattullo et al. (2014)	X		X	X	X	X					
Debernardis et al. (2014)	X		X	X	X						
Gattullo et al. (2015)	X		X	X	X	X					
Kruijff et al. (2019)	X		X	X	X				X		
Orlosky et al. (2013)		X		X						X	
Grout et al. (2015)		X		X							
Dingler et al. (2018)		X		X	X		X	X		X	
Hoffman et al. (2019)		X		X			X	X			
Tsunajima and Nishiuchi (2020)		X		X	X		X	X			X
Wang et al. (2020)		X		X	X		X	X			X
Wei et al. (2020)		X		X		X					X

Table 1 (continued)

	Aug-mented reality	Virtual reality	Optical see-through	Video see-through	Color and/or contrast	Text drawing style	Font type	Font size	Anchoring	Position	Text appearance and segmentation
Kojic et al. (2020)	X			X	X			X			
Erickson et al. (2020b)	X	X		X	X						
Buttner et al. (2020)	X	X		X	X				X		
Kobayashi et al. (2021)	X	X		X	X	X			X		
Shimizu et al. (2021)	X	X		X	X				X		
Rzayev et al. (2021)	X	X		X		X					X
Dewitz et al. (2021)	X	X		X		X				X	
Agić et al. (2022)	X	X		X	X		X			X	
Kojic et al. (2022)	X	X		X	X	X				X	X
Kobayashi et al. (2022)	X	X		X	X		X				
Total	28	17	28	21	24	9	6	14	7	15	8

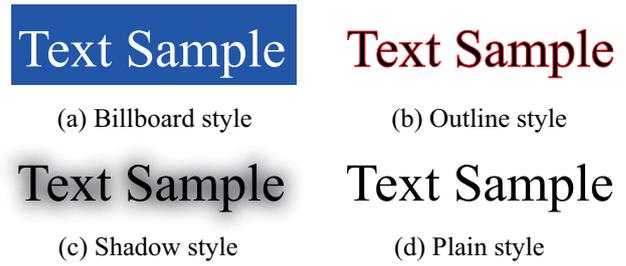


Fig. 2 Examples of text drawing styles: **a** billboard style, **b** outline style, **c** shadow style, and **d** plain style. (Created with Adobe Illustrator)

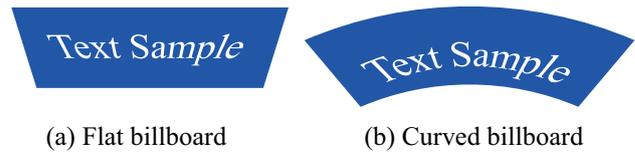


Fig. 3 Examples of flat billboard **a** and curved billboard **b** seen from the top. The horizontal black line helps to demonstrate the curve. (Created with Adobe Illustrator)

does not improve readability further (Gattullo et al. 2014, 2015).

Specific to OST displays – Falk et al. (2021) recommended the use of a solid billboard rather than a semi-transparent one. Regarding the outline and shadow styles, Gabbard et al. (2007) presented results that seem to indicate an equivalence between the two. Additionally, Gattullo et al. (2014, 2015) observed the inefficiency of the outline style even with the largest outline. Fiorentino et al. (2013) determined that combining the outline style and billboard style is inefficient since the two styles are better used alone than together.

From Grey Literature – Microsoft (2023a,b,c) recommends the use of a billboard to improve the readability of text. Google (2017, 2023) suggests the curvature of the UI elements, such as billboards, on a circle centered behind the user for a smooth presentation.

3.3 Color

In addition to color itself, we must begin by clarifying the concept of color polarity. Positive polarity (i.e., light mode) consists in using dark text on a light background, and negative polarity (i.e., dark mode) consists in using light text on a dark background (Buchner et al. 2009).

Specific to VST displays – Dingler et al. (2018) and Kojic et al. (2022) recommended respectively the negative and positive polarities both based on users’ preference. Erickson et al. (2020b) made an in-between by advising the negative polarity in case of dark lighting and, inversely, the positive

polarity in case of light lighting. They also suggest preferring a bright lighting (i.e., white = RGB(255, 255, 255)) rather than a dim lighting (i.e., white = RGB(25.5, 25.5, 25.5)). For the outline and plain style, Gattullo et al. (2014, 2015) recommended the positive polarity. Regarding color itself, Debernardis et al. (2014), Gattullo et al. (2015), and Kruijff et al. (2019) advised a white text on a blue billboard. If the color has a semantic, they suggested assigning the color to the billboard with a white text or a black text.

Specific to OST displays – The impact of background illuminance explained in Sect. 3.1 implies a preference for the negative polarity (Debernardis et al. 2014; Gattullo et al. 2014; Zhao et al. 2017; Kim et al. 2019; Erickson et al. 2021; Falk et al. 2021). A light text allows a more important contrast with the background and, inversely, a too light background does not allow any color to create the necessary contrast. Then, for a plain text style, the colors to be favored are for example cyan, green, and white (Gabbard et al. 2006, 2007; Fiorentino et al. 2013; Zhao et al. 2017). In the case of the billboard style, the contrast between the text and the box is more important than the one between the box and the background (Gabbard et al. 2007; Fiorentino et al. 2013; Debernardis et al. 2014; Rosilius et al. 2021). Inversely, for the outline and drop shadow styles, the contrast between the text and the drawing style is less important than the one between the drawing style and the background (Gabbard et al. 2007). In any case, black text must be avoided on a billboard since the black color is transparent on OST displays. The texture of the background will appear, making the readability correct only in the case of a sufficiently low background illuminance, i.e., when the background texture may be considered as black (Debernardis et al. 2014; Kim et al. 2019; Rosilius et al. 2021; Erickson et al. 2021). Regarding color itself, Debernardis et al. (2014), Gattullo et al. (2015), and Kruijff et al. (2019) advised a white text on a blue billboard. If the color has a semantic, they suggested assigning the color to the billboard with a white text or to the text with a white billboard.

Additionally, Gabbard et al. (2006, 2007) explored three algorithms to adapt the color to the background: the complement, the maximum HSV complement, and the maximum brightness contrast. The first is the inverse color in the RGB color space. The second is the saturated complement in the HSV color space. The third acts on the Y component (i.e., the spectral luminous efficiency function) of the color in the XYZ color space. The maximum brightness contrast performs better than the other two. However, their efficiency to improve the readability depends on the background. Thus, they were evaluated as globally less practical than the billboard style and the green plain text. In addition, they can be effective with the outline, drop shadow, and plain text styles, but not with the billboard style. As explained above, for the latter, it is the contrast between the text and the box, and not

between the text and the background, that is important. In the same way, Gabbard et al. (2007) and Fiorentino et al. (2013) reported that the billboard style must not be mixed with the outline style.

Finally, Gabbard et al. (2010), Sridharan et al. (2013) and Hincapie-Ramos (2014, 2015) studied the blending of colors. For their part, Gabbard et al. (2010) analyzed the hue movements of several colors depending on different backgrounds. They concluded that white backgrounds seem to have less impact on the hue than color backgrounds. Sridharan et al. (2013) and Hincapie-Ramos (2014, 2015) used similar analyses to propose a correction algorithm that defines the color to assign to the text to obtain a specific color once blending with the background. The algorithm consists in searching step-by-step for a color which will produce the desired color in a binned color gamut. Hincapie-Ramos (2014, 2015) proposed three implementations: the algorithm itself, the algorithm limited to the bins being a sufficient contrast with the background, and the algorithm associated with a billboard displayed only when the contrast is too low.

From Grey Literature – Microsoft (2023a,b,c) recommends the use of white text on dark or colored billboards. If the text must be black, they suggest using a bright billboard. Additionally, they advise the use of RGB(235, 235, 235) and RGB(16, 16, 16) respectively for the white and black color to prevent too bright experiences and to make the black visible on OST displays.

3.4 Text appearance and segmentation

Three reading methods will be discussed in this section: the paragraph presentation, the scrolling method, and the Rapid Serial Visual Presentation (RSVP). The former is the text displayed at once or per paragraph, but the studies rarely used more than one paragraph. The second is a presentation line-by-line. The latter is a method introduced by Forster (1970) that quickly presents the text word-by-word.

Specific to VST displays – Rzayev et al. (2021) advised the use of RSVP for short texts when the user must move in and interact with the world. In contrast, for a primary task without movement or a long text, they suggested the paragraph presentation. In addition, Kojic et al. (2022) recommended one or two maximum columns while Wang et al. (2020) recommended a minimal line spacing of 1.5. Finally, Tsunajima and Nishiuchi (2020) determined that the number of characters in one line is an important factor of readability, too many characters lead to the user being unable to read those on the sides. One effect of this is that more frequent head movements are required to scan the text, implying more fatigue and higher reading time. This issue was also observed by Dingler et al. (2018), who proposed 40 ± 6.6 characters in width and 7.3 ± 1.7 lines. Wei et al.

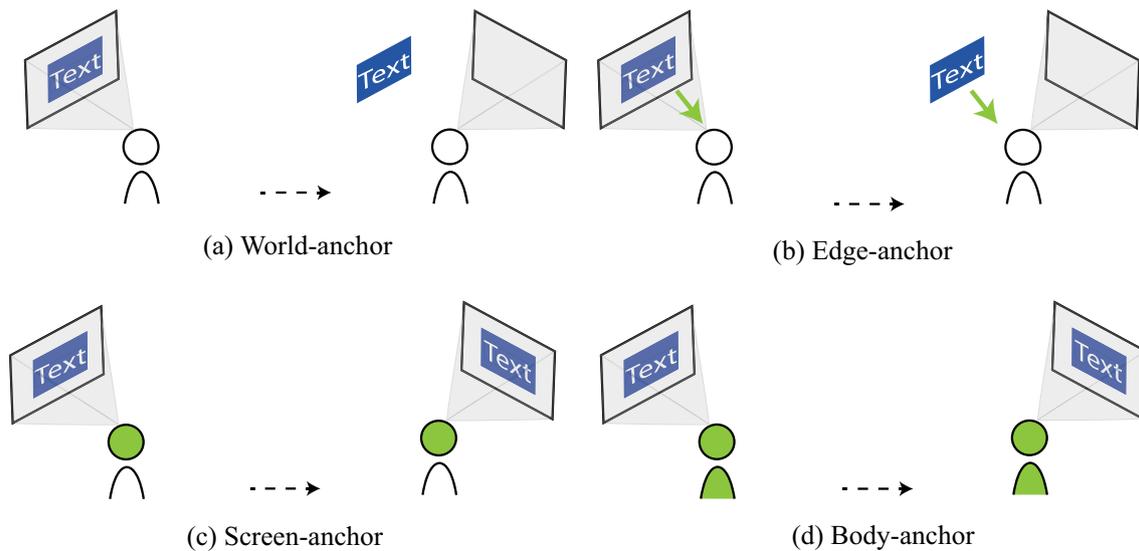


Fig. 4 Illustration of the four anchors: **a** world-anchor, **b** edge-anchor, **c** screen-anchor, and **d** body-anchor. The dotted arrow indicates that the user has moved. The green arrow in edge-anchor **b** indicates that

the text is always oriented toward the user. The green color on the user's body **c,d** indicates which part of the user's body the text follows. (Created with Adobe Illustrator)

(2020) fixed the field of view of the view box between 25.4° and 28.1° according to users' preference.

Specific to OST displays – Rzayev et al. (2018) evaluated two reading methods when walking and sitting. The results showed a preference for the scrolling method when walking and for the RSVP method when sitting. Falk et al. (2021) recommended to write the text in several short lines rather than a single long line.

From Grey Literature – Google (2023) and Microsoft (2023a,b,c) recommend using 2D text rather than 3D text because the extrusion deteriorates the readability of the text. Additionally, at a small font-size, they do not suggest thin or light weight text as it is more sensible to flickers and vibrations. Conversely, huge bold fonts do not allow proper discernment of the letters. In terms of letters recognition, Google (2023) advises ensuring sufficient space between them and to not use a halo that causes blur. Finally, Oculus (2023a,b) recommends left-aligning the text without justifying it, making text lines short, and subdividing text into sections.

3.5 Anchor

An anchor is the element to which another element is attached to define its position and rotation within the virtual world. We found four different anchors in the included papers: the world-anchor (see Fig. 4a), the edge-anchor (see Fig. 4b), the screen-anchor (see Fig. 4c), and the body-anchor (see Fig. 4d). The first fixes the position and rotation of the text at a specific location in the world. The

second is equivalent to the first, with the text oriented in permanence towards the user. The third fixes the position and rotation of the text relatively to the user's head. The last fixes the position and rotation of the text relatively to the user, but contrarily to the previous one, it is related to the body and not the head.

Specific to VST displays – Rzayev et al. (2021) recommended the edge-anchor or screen-anchor for the RSVP method and the edge-anchor or world-anchor for the paragraph presentation. However, Kobayashi et al. (2021) observed a preference for the world-anchor rather than the edge-anchor.

Specific to OST displays – Woodward et al. (2020) and Klose et al. (2019) recommended the world-anchor or body-anchor for non-essential information and everyday use, while the screen-anchor should be used for more important information or dual tasks (e.g., walking and reading simultaneously) needing permanent visual monitoring. Fukushima et al. (2020) suggested, for a walking user, the world-anchor when head movement is not critical and the screen-anchor when head movement is critical. Borg et al. (2015) explained that when walking, the human body activates some mechanisms to counterbalance movements such as those of the head and stabilize the user's view. According to them, it is a reason why the world-anchor should be favored over the screen-anchor. The latter generates retinal slip detrimental to the readability. Lee et al. (2023) recommended the body-anchor rather than the screen-anchor for textual notifications since it is better for comprehension and walking performance.

Table 2 Propositions of text placement by Google (2023) according to the type of text

Type of text	Distance to user	Viewing angle	Text state	User's state	Examples
“Text in hud”	0.5 – 1.5 m	Fixed	Stationary, moving	Still, moving	Time, notifications, text, updates, music track name
“Text for long reading”	0.5 – 5 m	Fixed, variable	Stationary	Still	Descriptions, articles, long-form reading
“Sticky info text”	0.5 – 5 m	Variable	Stationary	Still	Info over real-world objects
“Signage text”	Beyond 5 m	Variable	Stationary	Still, moving	Signages, billboards
“Responsive text”	0.5 – beyond 5 m	Variable	Stationary, moving	Still, moving	Navigation markers, responsive ads
“Ticker text”	0.5 – beyond 5 m	Fixed, variable	Moving	Still, moving	News, long-running info, ads

From *Grey Literature* – Google (2023) recommends the use of screen-anchor only for crucial information or short texts.

3.6 Position

Specific to VST displays – Shimizu et al. (2021) tested different methods to adapt the position of the text depending on the user point of attention. The results showed that playing on the vertical and horizontal axis is better than playing on the depth. In their paper, the vertical and horizontal axis represent the surface of a sphere centered on the user while depth corresponds to the sphere radius. Orlosky et al. (2013) observed a user tendency to place text 5.6° below screen center. Similarly, Dingler et al. (2018) observed a text position on a billboard at $-1 \pm 2.6^\circ$ from the horizontal. They also define two formulas for minimal and maximal distance vergence (see Eqs. 1 and 2) to be in acceptable range of vergence-accommodation conflict. The vergence-accommodation conflict occurs when the vergence distance (i.e., distance to the object) mismatches the accommodation distance (i.e., focal distance) such as on augmented and virtual reality headsets. d_v is the vergence distance, d_f is the focus distance, m_{near} , m_{far} , T_{near} , and T_{far} are the slopes and intercepts of the upper and lower bounds of vergence-accommodation comfort when considering vergence and focus on diopters.

Nevertheless, they observed that users often chose a distance beyond $d_{v,max}$ and then proposed a distance between 1 and 10 m with an ideal distance of 3 m on an HMD with a focal distance of 1.3 m. Kojic et al. (2022) recommended a depth of 6.5 m rather than 1.5 m for a panel of long text. Finally, Buttner et al. (2020) determined that a rotation of the text above 60° around the vertical axis (left, right) impacts significantly the readability.

$$d_v \min = \frac{d_f m_{near}}{1 - T_{near} d_f} \quad (1)$$

$$d_v \max = \frac{d_f m_{far}}{1 - T_{far} d_f} \quad (2)$$

Specific to OST displays – Research on OST displays is focused on the screen-anchor position. Klose et al. (2019) and Rzayev et al. (2018) determined that, for secondary information or too much complexity in the background, the text should be placed at the top-right, and, for primary information or dual-task needing a permanent monitoring, the text should be placed at the bottom-center. Lee et al. (2023) recommended placing notifications at 10° under the horizontal line. Koide et al. (2022) advised the use of a vertical alignment based on the head-movement for smaller depth (0.45 m) when walking. However, they observed that its utility decreased with increasing depth (0.75 m and 1.5 m). For the depth, Woodward et al. (2020) advised against a too long depth (0.5 m) because the information is then considered secondary and does not need to remain fixed to the user view. So, it is recommended to use the world-anchor in this case. Gabbard et al. (2019) recommended a small distance (e.g., 2 m) to prevent context and focal switching. Arefin et al. (2022) concluded that context switching had no effect on task performance but increases fatigue.

From *Grey Literature* – Google (2023) defines the comfortable area of view between $30\text{--}35^\circ$ and 120° when rotating the head. The center of the view is $\pm 6^\circ$ under the horizontal line. In Table 2, they propose the placement of text according to the type of information. The viewing angle defines whether the text is always facing the user even if they move around it. The text state determines whether the text moves with the user. The user's state indicates whether the user remains still or is moving. The distance to the user is also indicated in the table. On this topic, Microsoft (2023a,b,c) recommends a minimum of 0.4 m and a comfortable area between 1.25 m and 5 m. Oculus (2023a,b) indicates that an arm length of depth is sufficient.

3.7 Fonts-type

Specific to VST displays – Dingler et al. (2018) recommended the Arial (sans-serif) font for English text rather than Times New Roman (serif) font. Kobayashi et al. (2021, 2022) focused on Japanese characters. In their first work, they observed that Yu Mincho (serif) font is significantly

better than Yu Gothic (sans-serif) font. Their later work highlighted a tendency for the Antigothic font family to be more readable and less fatiguing than the Mincho and Gothic font families.

Specific to OST displays – Zhao et al. (2017) suggested the use of Verdana (sans-serif) font compared to Times New Roman (serif) for people with moderate or severe low vision.

From Grey Literature – Oculus (2023a,b) indicates that sans-serif fonts are more legible than serif or stylized fonts. However, Google (2023) agrees with that statement only for high-contrast serif typefaces. Moreover, they recommend low to moderate typefaces that require fewer pixels for presenting details. In general, they recommend using fonts with wider letters and that are not condensed due to the important impact of the depth on the readability. For that, they advise against Helvetica and Univers fonts. Yet, Microsoft (2023a,b,c) recommends the former with the Segoe UI and Arial fonts. Additionally, Microsoft (2023a,b,c) recommends limiting to two fonts family. Finally, Oculus (2023a,b) suggest using OpenDyslexic, Arial or Comic Sans fonts.

3.8 Fonts-size

Whether on OST or VST displays, Kruijff et al. (2019) observed that users preferred a font-size larger for text at the edge of the field of view than for text at the center. To describe writing sizes, we have converted all paper suggestions to the angular unit named Distance-Independent Millimeters (dmm) (Google 2017). dmm consists of x millimeter at one meter of distance where x is the coefficient of the unit [e.g., 1 dmm = 1 mm at one meter or 2 mm at two meters (Google 2017)].

Specific to VST displays – In regard to the Latin alphabet, Dingler et al. (2018) recommended a font-size of $32 \text{ dmm} \pm 11 \text{ dmm}$ for a capital letter and $41 \text{ dmm} \pm 14 \text{ dmm}$ for a body text. Similarly, Hoffman et al. (2019) suggested a comfortable font-size of 26 dmm for label (capitalized) and body text. Additionally, they observed a correlation with the contrast of the text. The more the contrast decreases under the minimal value, the bigger the font-size must be to compensate. As for Kojic et al. (2020), they obtained similar results for short texts (2 words) with a font-size of 30 dmm, but for longer texts between 21 and 51 words, the needed font-size is reduced to 17 dmm. The explanation given by the authors was that participants consider the short text as a title and, therefore, gave it a larger preferred size. Dewitz et al. (2021) aligned themselves with these results, while they found a comfortable font-size of 15 dmm. Additionally, Agic et al. (2022) proposed a font-size of 32-40 pt at 5 m, 36-56 pt at 10 m, and 56-68 pt at 15 m on an HTC Vive.

Considering logographic alphabet, Kobayashi et al. (2021) recommended a visual angle between 19 and 21 dmm

Table 3 Font-sizes recommended by Google (2023) depending on the type of text

Type of text	Weight	Font-size (dmm)
Headline	Regular	40
Title	Medium	32
Subheading	Regular	28
Body 2	Medium	24
Body 1	Regular	24
Caption	Regular	20
Button	Medium	24

while Wang et al. (2020) proposed a font-size of 12 pt or more at 0.5 m on a Samsung Gear VR.

Specific to OST displays – Erickson et al. (2021) observed a positive correlation between the depth and the font-size. Borg et al. (2015) recommended a font-size between 5 and 17 dmm when standing. If the user must walk, the lower limit becomes 9 dmm. They also observed that the font-size of screen-anchor text was more affected at the extremity of the range than world-anchor text. Renkewitz et al. (2008) observed a minimal font-size of 16 dmm to reach a maximal 2-s recognition time. They also concluded that recommendations on font-size for desktop may be applied to HMDs. Gabbard et al. (2019) observed a minimal font-size of 5 dmm for text at 6 m of depth. Finally, Zhao et al. (2017) recommended a minimal font-size of 100 px at 3 m on an Epson BT-200 for people with moderate or severe low vision.

Considering logographic alphabets, Chang et al. (2019) observed a mean font-size of 4 dmm, but to reach the 99th percentile legibility threshold, the font-size must be to 7 dmm. However, they observed, on the one hand, that the more strokes in the character, the more the font-size must be large, and, on the other hand, the more components in the character (i.e., separated groups of strokes), the easier to read the character is.

From Grey Literature – First, Microsoft (2023a,b,c) indicates not to rely on computer font-size because of the impact of depth. Second, Oculus (2023a,b) suggests beginning with a font-size of 10% of the screen and adapt to the requirements. Microsoft (2023a,b,c) defines a comfortable font-size between 11 and 14 dmm at 0.45 m and 11 dmm and 13 dmm at 2 m. Finally, the font-sizes recommended by Google (2023) are presented in Table 3.

4 Results by aggregation of text parameters

We aggregated the parameters that were intrinsically related. On the one hand, these are the text drawing style and the color (see Fig. 5), and, on the other hand, these are the

anchor and the position (see Fig. 6). As for the other parameters, they do not depend on a sequence of decisions to be taken, are not sufficiently studied, or are subject to too many contradictions in the literature to propose a clear decision tree. In addition, while it is true that contrast and font-size are related, as observed by Gattullo et al. (2014), it is a question of balancing the two to reach a sufficient readability threshold. Similarly, font-size and depth are related, since the font-size is expressed in angular units, it will automatically adapt to the choice made for the distance between the user and the text.

The two decision trees are represented in Figs. 5 and 6. They start with the green node “Start” and finish with the green nodes containing the list of guidelines for the choices made according to the route. The yellow blocks represent decisions, while the blue blocks represent major choices to respect. Numbers in square brackets refer to a set of documents that are the source of the guideline. In Fig. 5, the choice of device is the first question, as it determines the rest of the process for the text drawing style and color. On the contrary, in Fig. 6, we hypothesized that results for both types of display may be merged since it does not seem to depend on device-related criteria. The decision trees illustrate that there is still room for further research to improve the guidelines. For example, in Fig. 5, we observed, on a VST display, that for a billboard without semantic on the color (see bottom-right part of the figure), the designer must choose which literature they follow among three exclusive possibilities: (1) (Debernardis et al. 2014; Gattullo et al. 2015; Dingler et al. 2018; Kruijff et al. 2019), (2) (Erickson et al. 2020b), and (3) (Kojic et al. 2022).

5 Discussion

In this section, we provide a detailed discussion on the three research questions. For each, we present the results obtained from our literature review and we elaborate on future research avenues these results open.

5.1 Research question 1: Text parameters

The first research question is: “What are the different text parameters and how to tune them to maximize the readability of a text?”. We identified seven parameters in the literature: text drawing style, color, anchor, position, font-type, font-size, and text appearance and segmentation. We observed that the text drawing style and the color, as well as the anchor and the position were respectively intrinsically related. Furthermore, we also found a relationship of balance between the contrast and the font-size. The lower the contrast, the larger the font-size needs to be. In addition, we

identified an intrinsic relation between the font-size and the depth, but the use of angular units for the first allows them to be treated separately.

Concerning the drawing style of the text, a consensus was reached in the literature to favor the use of billboards (i.e., a panel box behind the text) except when the occlusion is a problem for the user. In that case, authors recommend the use of a plain text style, since the outline style and shadow style do not present better performance. On Video See-Through (VST) displays, a small outline can be added to improve readability. Regarding the shadow style, it was studied once on Optical See-Through (OST) displays. In terms of color, on OST displays, the light-additive property implies the use of negative polarity. In comparison, on VST displays, the literature is more divided, advising both positive and negative polarity. However, for both displays, the colors that stand out most are white for the text and blue for the billboard. In case color has semantics, it is recommended to assign the semantically meaningful color to the billboard.

Regarding the anchor of the text, research on VST displays is more concerned with the way the text is presented, while research on OST displays is more concerned with the type of text and the simultaneous movement of the user. Generally, when it comes to short text such as notifications or dashboard-type information, it is recommended to use a screen-anchor approach. For long text or text associated with objects, previous works recommend the use of world-anchor or edge-anchor texts. Nevertheless, some studies demonstrated that the human body has a system for balancing eyesight when walking, which conflicts with the screen-anchor. In terms of position, the center of the view is located at $\pm 6^\circ$ below the horizontal line. Text is also recommended at the bottom of the view when attached to the user's view. Additionally, most of the text should be displayed at minimum 0.4 – 0.5 m and maximum 5–10 m. Once again, the type of text will influence the depth at which it is placed. In Sect. 4, we hypothesized that the results from both types of display may be merged, but this should be verified in future works.

Considering the font-type, the White Literature seems to indicate that fonts with too much detail, such as the serif family, should be avoided. However, while some Grey Literature sources support this idea, others reject it. Some fonts have been recommended, but again the literature does not always reach consensus on this. Yet, the Arial font remains a standard choice. In terms of font size, all surveyed studies found different numbers. Recommendations vary between 5 and 17 dmm, with some going as high as 40 dmm. This recommendation applies to both Latin and logographic alphabets. It must however be noted that in this survey we presented only comfortable thresholds instead of minimal font sizes. Additionally, some works give their recommendations in a unit that depends on the pixel size, and thus on the

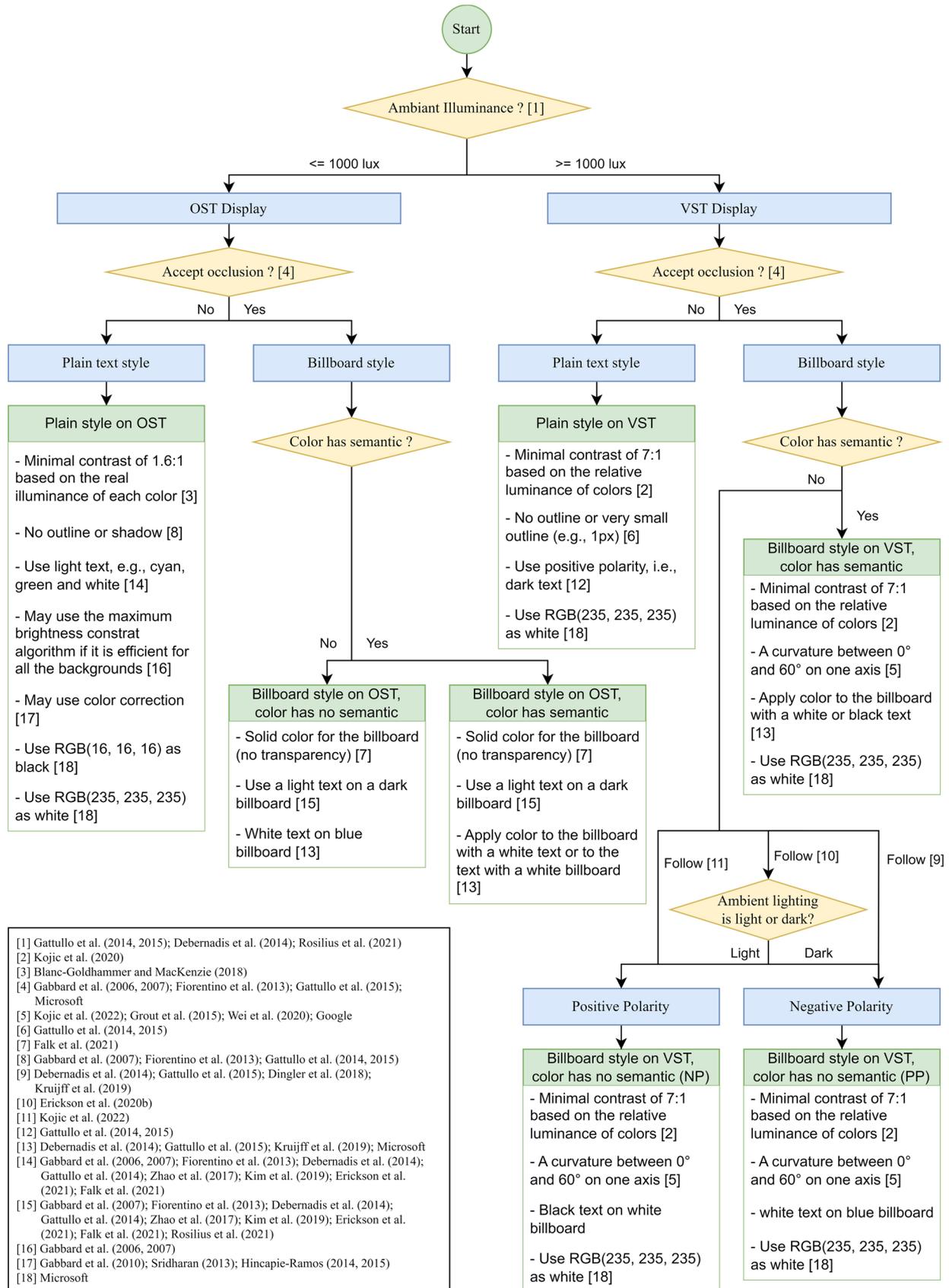


Fig. 5 Decision-tree to select the right text drawing style and color for a text. (Created with draw.io)

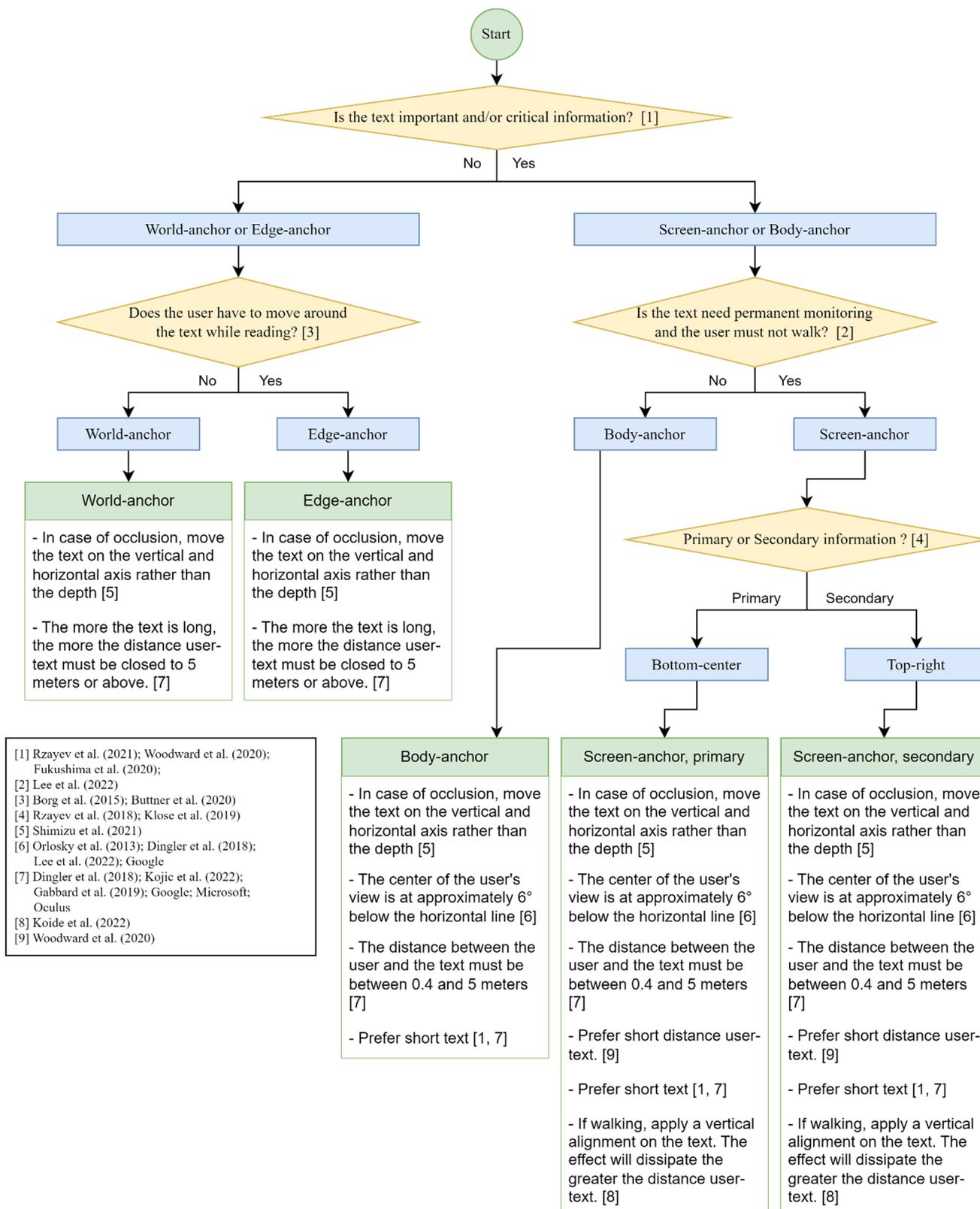


Fig. 6 Decision-tree to select the right anchor and position for a text. (Created with draw.io)

device used. These are complicated to compare with other works that use angular units. Last, it is not clear how distance influences the minimum and maximum bounds, even if these are expressed in angular units. Do they remain fixed, or are they dependent on a distance-based function? The vergence-accommodation conflict could potentially play a role in favor of the second proposal, but this remains to be verified in future research.

We grouped under the text appearance and segmentation label all guidelines that have not been sufficiently studied to have their own category, such as line spacing, weight, text length, number of columns, text alignment, and presentation mode. The extent of their impact on readability is unclear and needs further study. Classifying parameters by order of importance could also help designers to better control their experiences, while letting them indulge their artistic creativity. From the results, we can already conclude that contrast, depth, and font size are critically important, but the question remains open for other parameters, calling for novel research in these still shady areas.

In view of the answer to the first research question, we recommend focusing future research efforts on four different avenues: (1) better define the rules around parameters under the text appearance and segmentation label, (2) clarify the range of optimal text size, (3) define which results obtained on VST (resp. OST) displays can also be applied on OST (resp. VST) displays without repeating the study, and (4) define a taxonomy of text usage and associate readability rules to each.

5.2 Research question 2: Contextual constraints

The second research question is: “What are the contextual constraints to consider when designing an application including texts?”. We determined six constraints: the type of display, the background texture, the background illuminance, the user’s task, the dual task such as walking, and the user’s body. However, the type of display and the background illuminance are intrinsically related.

As already discussed, the type of display impacts the way the device renders the text, but also on which devices the guidelines were tested. Gattullo et al. (2014, 2015) and Debernardis et al. (2014) recommended using OST displays with an ambient illuminance of less than 1,000 lx and VST displays with an ambient illuminance of more than 1,000 lx. In the case of OST displays, the light-additive property does not allow the device to be exposed to too much light. In the case of VST displays, the cameras normalize brightness, making dark environments too dark to use the headset. Nevertheless, this does not prevent them from being used below 1,000 lx, but there is a preference for OST displays. Additionally, too low light (10–12 lx) is not suitable for OST displays (Kim et al. 2019). Concerning the background texture,

it impacts the readability of text only when the contrast is low or when the texture is complex (Leykin et al. 2004; Gabbard et al. 2006). On one hand, as with conventional devices (i.e., computer or mobile), the color difference between the text and the background on which it is read must be sufficient to determine the outline of the letters. On another hand, Gabbard et al. (2006) determined that it was the complexity of the background in the area close to the text that had an impact on its reading, rather than the complexity of the background in general. In their experience, they observed that participants obtained good results with the brick wall because they place the text on a single brick that presents a degree of homogeneity. Therefore, the readability of a text depends on the disturbance of the background texture close to the text, generating variation in contrast between the two. However, the designer can have a control on that using either dynamic algorithms such as the ones proposed by Gabbard et al. (2010), Sridharan et al. (2013), and Hincapie-Ramos (2014, 2015), or classifiers (Scharff et al. 1999; Leykin and Tuceryan 2004; Tanaka et al. 2008; Manghisi et al. 2017).

Regarding the user’s task and dual task (e.g., walking and reading simultaneously), there is no doubt about the impact on text readability. However, there is a need for more in-depth analysis, especially on the user’s task. At present, the only interest is in knowing whether the occlusion will be a problem for the user. For the dual task, it was demonstrated that the more the dual task needs attention, the more the readability decreases (Klose et al. 2019). Additionally, Wei et al. (2020) identified a trade-off between the ease of reading and an immersive reading experience. In this respect, the task and its context can help define the level of immersion required.

Last, similarly to desktop UIs, the user’s body has a significant impact. First, when walking, the view is automatically balanced to avoid too much jerking that is incompatible with certain anchoring methods. Second, the vergence-accommodation conflict plays an important role in the reading of text. Dingler et al. (2018) offer formulas for determining a range of acceptability, but the problem is to find solutions that play on the focus of the device. Therefore, in future work, it would be interesting to look at solutions based on more than just the text itself. Apart from this, several studies have reported that the user does not always make appropriate choices in relation to measures such as accuracy and reading speed (Gattullo et al. 2015; Kim et al. 2019). Additionally, Zhao et al. (2017) concluded that “if a sighted person can use the AR glasses while walking, it is also feasible for a low vision person to use the glasses while walking”.

In regard to the answer to the second research question, we recommend pursuing future research efforts on two main avenues. First, the hardware development of displays should eventually solve several of the above-mentioned problems. For instance, successfully combining the strengths of both

types of devices, either with special lenses or with a transparent VST screen. Secondly, the use of algorithms combined with sensor data could help reduce environmental and/or user's metabolic problems such as those trying to counterbalance the walk or the brightness of the background.

5.3 Research question 3: Gap between white and grey literature

The third research question is: "What is the gap between the White and Grey Literature in the guidelines proposed on text readability?". Among the sites we visited, only three device manufacturers have guidelines on text readability in their documentation. On the color, contrast, text drawing style, font-type, and font-size, the literatures are similar, but for the rest, they tend to talk about one side but not the other, and vice versa. This observation strengthens the case for including Grey Literature in a literature review like ours. Additionally, it seems that the White Literature is more comprehensive, showing signs of difficulty in transferring information from the scientific side to the industrial side. However, we observed that Google (2023) cites the works of Gabbard et al. (2006) and Gattullo et al. (2015). Their documentation also appears to be written by, or at least with, an independent researcher (Niteesh 2023). Just how much of this documentation is based on scientific sources is unclear. Perhaps, if it is not already the case, documentation should be written more often in collaboration with researchers, and thus, promote popularization through media other than scientific books, articles, and conferences (or at least reference these works in documentation). Another approach would be to offer works like this one, and above all continuing to keep it up to date regularly on a platform, to facilitate access to scientific research in industries and, possibly, directly to the public. However, the question of who would be responsible for such a project remains difficult to answer. Besides, we observed a more rigorous classification of text from the Grey Literature than the White Literature. As already discussed in Sect 5.1, , it would be interesting to use such a classification in the future, to indicate more clearly what type of information is being studied. In fact, it has been shown that requirements differ according to the role of the text.

5.4 Limitations

Variability in experimental protocols makes it difficult to collate and compare results from dozens of papers, since the number of parameters involved is too large. We decided not to apply quality criteria, even if some papers contradict themselves. In such cases, we chose to take the version that was presented in their discussion. It would be interesting in future work to generalize and confirm the results obtained to date. Even more so as the question arises as to whether

some of the results achieved on older headsets are still relevant today, given the technological advances of today's headsets. Similarly, the question will arise in the future with recently published papers. Eventually, we will need to be able to determine more precisely the impact of hardware on results. Additionally, the nomenclature associated with text readability domains is diverse and varied, a problem also highlighted by Erickson et al. (2020b).

Furthermore, we limited our scope to papers that studied Head-Mounted Displays (HMDs). Nonetheless, some works on other devices, such as Head-Up Displays (HUDs), present results that may be applied to HMDs. According to Renkewitz et al (2008), this is the case for font size guidelines on desktop systems, but it is a conclusion that has not been shared by all works dealing with this text parameter. It is also possible to include all the literature on classifiers that try to determine the readability of text based on images. If we study the criteria used by the best-performing models, we may discover or confirm results obtained via more conventional approaches, such as the work included in this survey.

Finally, we have included results from analyses such as reading speed or number of errors, as well as results from user preferences. However, for the latter, it is well known that many social, cultural, and other factors have a major impact on results. Additionally, Gattullo et al. (2015) and Kim et al. (2019) demonstrated that users do not always make the choice of performance. A whole area of research remains to be carried out to verify the results and determine the relationship between readability and enjoyment.

6 Conclusion

We conducted a multivocal literature review to identify the list of text parameters that can be tuned to improve text readability, as well as the associated contextual constraints, in order to propose an up-to-date state of knowledge and guidelines. We adopted a multivocal protocol to include the Grey Literature of high control and credibility. In total, we found seven text parameters and six contextual constraints. The most important ones are the contrast (i.e., related to the text drawing style and color), the font-size, the anchor, and the position. They are impacted by the background, the role of the text and the user's body, such as the vergence-accommodation conflict. Nevertheless, the guidelines pertaining to those parameters seem complete and accepted by the community, except for some minor points. The same cannot be said for the other parameters, which are either under-researched or the subject of contradictions among scholars and/or practitioners. In addition, this research also highlighted a lack of communication of guidelines from White Literature to Grey Literature. All in all, this survey has gathered numerous guidelines for text design on OST

and VST HMDs. They have been brought together to form a guide, assisting designers in the choice of the best solution for the context in which text is to be used. This will be of use both for researchers wishing to extend this research, and for designers wishing to ensure the readability of their texts in their applications. Finally, we conclude with the proposition of future research directions to extend the results and to promote them outside scientific publications.

Acknowledgements We would like to thank the project nb. 8482 “Région Wallonne (RW) Mecatech Federated Learning and Augmented Reality for Advanced Control Centers (FLARACC)” for enabling and funding this research.

Author contributions Maxime Cauz conducted the study and wrote the paper. Antoine Clarinval and Bruno Dumas supervised the process and reviewed the paper.

Funding Project nb. 8482 “Région Wallone (RW) Mecatech Federated Learning and Augmented Reality for Advanced Control Centers (FLARACC)”.

Région Wallone, Belgium, 8482

Availability of data and materials It is not applicable to this survey.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical Approval It is not applicable to this survey.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adams RJ, Smart P, Huff AS (2017) Shades of grey: Guidelines for working with the grey literature in systematic reviews for management and organizational studies. *Int J Manag Rev* 19(4):432–454. <https://doi.org/10.1111/ijmr.12102>
- Agić A, Mandić L, Loknar NS (2022) Legibility of typefaces and preferences of text/background color variations in virtual environment. In: Proceedings of the 11th international symposium on graphic engineering and design (GRID), pp 835–841. <https://doi.org/10.24867/grid-2022-p92>
- Arefin MS, Philips N, Plopski A, Gabbard JL, Swan JE (2022) The effect of context switching, focal switching distance, binocular and monocular viewing, and transient focal blur on human performance in optical see-through augmented reality. *IEEE Trans Visual Comput Graphics* 28(5):2014–2025. <https://doi.org/10.1109/tvcg.2022.3150503>
- Azuma RT (1997) A survey of augmented reality. *Presence: Teleoperators Virtual Environ* 6(4):355–385. <https://doi.org/10.1162/pres.1997.6.4.355>
- Blanc-Goldhammer DR, MacKenzie KJ (2018) The effects of natural scene statistics on text readability in additive displays. *Proc Human Fact Ergon Soc Annual Meeting* 62(1):1281–1285. <https://doi.org/10.1177/1541931218621294>
- Borg O, Casanova R, Bootsma RJ (2015) Reading from a head-fixed display during walking: Adverse of gaze stabilization mechanisms. *PLoS ONE* 10(6):e0129902. <https://doi.org/10.1371/journal.pone.0129902>
- Buchner A, Mayr S, Brandt M (2009) The advantage of positive text-background polarity is due to high display luminance. *Ergonomics* 52(7):882–886. <https://doi.org/10.1080/00140130802641635>
- Buttner A, Grunvogel SM, Fuhrmann A (2020) The influence of text rotation, font and distance on legibility in VR. In: Proceedings of the IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW), pp 662–663. <https://doi.org/10.1109/vrw50115.2020.00182>
- Cárdenas-Robledo LA, Hernández-Uribe Ó, Reta C, Cantoral-Ceballos JA (2022) Extended reality applications in industry 4.0. – A systematic literature review. *Telemat Inform* 73:101863. <https://doi.org/10.1016/j.tele.2022.101863>
- Chang, C-C, Liu C-J, Liu K-H (2019) The legibility threshold of Chinese characters on the augmented reality display - A preliminary study. In: Proceedings of the 50th nordic ergonomics and human factors society (NES) conference, pp 69–72.
- Debernardis S, Fiorentino M, Gattullo M, Monno G, Uva AE (2014) Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices. *IEEE Trans Visual Comput Graphics* 20(1):125–139. <https://doi.org/10.1109/tvcg.2013.86>
- Dewitz B, Geiger C, Steinicke F (2021) Virtual Visus-Vision Acuity and Text Legibility in Virtual Environments. *GI VR/AR Workshop*. https://doi.org/10.18420/vrar2021_5
- Dingler T, Kunze K, Outram B (2018) VR Reading UIs: Assessing Text Parameters for Reading in VR. In: Extended abstracts of the 2018 CHI conference on human factors in computing systems (CHI EA), pp 1–6. <https://doi.org/10.1145/3170427.3188695>
- Erickson A, Kim K, Bruder G, Welch G (2020a) A Review of Visual Perception Research in Optical See-Through Augmented Reality. In: Proceedings of the international conference on artificial reality and telepresence, eurographics symposium on virtual environments (ICAT-EGVE). <https://doi.org/10.2312/egve.20201256>
- Erickson A, Kim K, Bruder G, Welch GF (2020b) Effects of Dark Mode Graphics on Visual Acuity and Fatigue with Virtual Reality Head-Mounted Displays. In: Proceedings of the IEEE conference on virtual reality and 3D user interfaces (VR), pp. 434–442. <https://doi.org/10.1109/vr46266.2020.00064>
- Erickson A, Kim K, Lambert A, Bruder G, Browne MP, Welch GF (2021) An extended analysis on the benefits of dark mode user interfaces in optical see-through head-mounted displays. *ACM Trans Appl Perception* 18(3):1–22. <https://doi.org/10.1145/3456874>
- Falk J, Eksvard S, Schenkman B, Andren B, Brunnstrom K (2021) Legibility and readability in Augmented Reality. In: Proceeding of the 13th international conference on quality of multimedia experience (QoMEX), pp 231–236. <https://doi.org/10.1109/qomex51781.2021.9465455>
- Fiorentino M, Debernardis S, Uva AE, Monno G (2013) Augmented reality text style readability with see-through head-mounted displays in industrial context. *Presence: Teleoperators Virtual Environ* 22(2):171–190. https://doi.org/10.1162/pres_a_00146

- Fukushima S, Hamada T, Hautasaari A (2020) Comparing world and screen coordinate systems in optical see-through head-mounted displays for text readability while walking. In: Proceedings of the IEEE international symposium on mixed and augmented reality (ISMAR), pp 649–658. <https://doi.org/10.1109/ismar50242.2020.00093>
- Gabbard JL, Swan JE, Hix D (2006) The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. Presence: Teleoperators Virtual Environ 15(1):16–32. <https://doi.org/10.1162/pres.2006.15.1.16>
- Gabbard JL, Mehra DG, Swan JE (2019) Effects of AR display context switching and focal distance switching on human performance. IEEE Trans Visual Comput Graphics 25(6):2228–2241. <https://doi.org/10.1109/tvcg.2018.2832633>
- Gabbard JL, Swan JE, Hix D, Kim S-J, Fitch G (2007) Active text drawing styles for outdoor augmented reality: A user-based study and design implications. In: Proceedings of the IEEE virtual reality conference, pp 35–42. <https://doi.org/10.1109/vr.2007.352461>
- Gabbard JL, Swan JE, Zedlitz J, Winchester WW (2010) More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In: Proceedings of the IEEE virtual reality conference (VR), pp. 79–86. <https://doi.org/10.1109/vr.2010.5444808>
- Garousi V, Felderer M, Mäntylä MV (2019) Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. Inf Softw Technol 106:101–121. <https://doi.org/10.1016/j.infsof.2018.09.006>
- Gattullo M, Uva AE, Fiorentino M, Monno G (2014) Effect of text outline and contrast polarity on AR text readability in industrial lighting. IEEE Trans Visual Comput Graphics 21(5):638–651. <https://doi.org/10.1109/tvcg.2014.2385056>
- Gattullo M, Uva AE, Fiorentino M, Gabbard JL (2015) Legibility in industrial AR: Text style, color coding, and illuminance. IEEE Comput Graphics Appl 35(2):52–61. <https://doi.org/10.1109/mcg.2015.36>
- Gattullo M, Evangelista A, Uva AE, Fiorentino M, Gabbard JL (2022) What, how, and why are visual assets used in industrial augmented reality? A systematic review and classification in maintenance, assembly, and training (from 1997 to 2019). IEEE Trans Visual Comput Graphics 28(2):1443–1456. <https://doi.org/10.1109/tvcg.2020.3014614>
- Google for Developers (2017) Designing screen interfaces for VR (Google I/O '17). https://www.youtube.com/watch?v=ES9jArHRFHQ&ab_channel=GoogleforDevelopers. [Accessed 19 June 2023]
- Google Fonts (2023) Using type in AR & VR. Google fonts documentation. https://fonts.google.com/knowledge/using_type_in_ar_and_vr. [Accessed 19 June 2023]
- Grout C, Rogers W, Apperley M, Jones S (2015) Reading text in an immersive head-mounted display: An investigation into displaying desktop interfaces in a 3D virtual environment. In: Proceedings of the 15th New Zealand conference on human-computer interaction, pp 9–16. <https://doi.org/10.1145/2808047.2808055>
- Hincapie-Ramos JD, Ivanchuk L, Sridharan SK, Irani P (2014) SmartColor: Real-time color correction and contrast for optical see-through head-mounted displays. In: Proceedings of the IEEE international symposium on mixed and augmented reality (ISMAR), pp 187–194. <https://doi.org/10.1109/ismar.2014.6948426>
- Hincapie-Ramos JD, Ivanchuk L, Sridharan SK, Irani PP (2015) SmartColor: Real-time color and contrast correction for optical see-through head-mounted displays. IEEE Trans Visual Comput Graphics 21(12):1336–1348. <https://doi.org/10.1109/tvcg.2015.2450745>
- Hoffman DM, McKenzie C, Koprowski B, Iqbal A, Balram N (2019) Aligning content rendering resolution and feature size with display capability in near-eye display systems. J Soc Inform Display 27(4):207–222. <https://doi.org/10.1002/jsid.765>
- Kaufeld M, Mundt M, Forst S, Hecht H (2022) Optical see-through augmented reality can induce severe motion sickness. Displays 74:102283. <https://doi.org/10.1016/j.displa.2022.102283>
- Kim A-Y, Song H-J, Park S-B, Lee S-J (2014) Device-dependent readability for improved text understanding. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP), pp 1396–104. <https://doi.org/10.3115/v1/d14-1146>
- Kim K, Erickson A, Lambert A, Bruder G, Welch G (2019) Effects of dark mode on visual fatigue and acuity in optical see-through head-mounted displays. In: 19th Symposium on spatial user interaction (SUI), pp 1–9. <https://doi.org/10.1145/3357251.3357584>
- Kitchenham BA, Budgen D, Brereton P (2015) Evidence-based software engineering and systematic reviews. CRC Press, New York. <https://doi.org/10.1201/b19467>
- Klose EM, Mack NA, Hegenberg J, Schmidt L (2019) Text presentation for augmented reality applications in dual-task situations. In: Proceedings of the IEEE conference on virtual reality and 3D user interfaces (VR), pp 636–644. <https://doi.org/10.1109/vr.2019.8797992>
- Kobayashi S, Kanari K, Sato M (2021) An examination of view-settings for long texts in VR reading. In: Special interest group on computer graphics and interactive techniques conference (SIGGRAPH) posters, article 10, pp 1–2. <https://doi.org/10.1145/3450618.3469164>
- Kobayashi S, Kanari K, Sato M (2022) Effects of font type and weight on reading in VR. In: 22' SIGGRAPH Asia 2022 posters, pp 1–2. <https://doi.org/10.1145/3550082.3564182>
- Koide H, Kanari K, Sato M (2022) An examination on reduction of displayed character shake while walking in place with AR glasses. In: Proceedings of the IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW), pp 786–787. <https://doi.org/10.1109/vrw55335.2022.00244>
- Kojic T, Ali D, Greinacher R, Moller S, Voigt-Antons J-N (2020) User experience of reading in virtual reality — finding values for text distance, size and contrast. In: Proceedings of the 12th international conference on quality of multimedia experience (QoMEX), pp 1–6. <https://doi.org/10.1109/qomex48832.2020.9123091>
- Kojic T, Vergari M, Möller S, Voigt-Antons J-N (2022) Assessing user experience of text readability with eye tracking in virtual reality. In: Proceedings of the international conference on human-computer interaction (HCI) 2022: Virtual, augmented and mixed reality: Design and development, pp 199–211. https://doi.org/10.1007/978-3-031-05939-1_13
- Korkut EH, Surer E (2023) Visualization in virtual reality: a systematic review. Virtual Reality 27(2):1447–1480. <https://doi.org/10.1007/s10055-023-00753-8>
- Kruijff E, Orlosky J, Kishishita N, Trepkowski C, Kiyokawa K (2019) The influence of label design on search performance and noticeability in wide field of view augmented reality displays. IEEE Trans Visual Comput Graphics 25(9):2821–2837. <https://doi.org/10.1109/tvcg.2018.2854737>
- Lawrence A., Houghton J., Thomas J., Weldon P. (2014) Where is the evidence? Realising the value of grey literature for public policy & practice, A discussion paper. Swinburne Institute for Social Research.
- Lee H, Bang S, Woo W (2023) Effects of coordinate system and position of AR notification while walking. Virtual Reality 27(2):829–848. <https://doi.org/10.1007/s10055-022-00693-9>
- Leykin A, Tuceryan M (2004) Determining text readability over textured backgrounds in augmented reality systems. Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry (VRCAI), pp 436–439. <https://doi.org/10.1145/1044588.1044683>

- Liberatore MJ, Wagner WP (2021) Virtual, mixed, and augmented reality: a systematic review for immersive systems research. *Virtual Reality* 25(3):773–799. <https://doi.org/10.1007/s10055-020-00492-0>
- Manghisi VM, Gattullo M, Fiorentino M, Uva AE, Marino F, Bevilacqua V, Monno G (2017) Predicting text legibility over textured digital backgrounds for a monocular optical see-through display. *Presence* 26(1):1–15. https://doi.org/10.1162/pres_a_00285
- Microsoft (2023a) Typography. Microsoft Learn Documentation. <https://learn.microsoft.com/en-us/windows/mixed-reality/design/typography>. Accessed 19 June 2023
- Microsoft (2023b) Color, light, and materials. Microsoft Learn Documentation. <https://learn.microsoft.com/en-us/windows/mixed-reality/design/color-light-and-materials>. Accessed 19 June 2023
- Microsoft (2023c) Comfort. <https://learn.microsoft.com/en-us/windows/mixed-reality/design/comfort>. Accessed 19 June 2023
- Munafo J, Diedrick M, Stoffregen TA (2016) The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Exp Brain Res* 235(3):889–901. <https://doi.org/10.1007/s00221-016-4846-7>
- Niteesh Y (2023) Website for their independent research. <https://niteeshyadav.com/> Accessed 19 June 2023
- Oculus Developers (2023a) VR Accessibility Design: User Experience (UX) and User Interface (UI). <https://developer.oculus.com/resources/design-accessible-vr-ui-ux/#text>. Accessed 19 June 2023
- Oculus Developers (2023b) VR Accessibility Design: Captions and Subtitles. <https://developer.oculus.com/resources/design-accessible-vr-captions/>. Accessed 19 June 2023
- Orlosky J, Kiyokawa K, Takemura H (2013) Towards intelligent view management: A study of manual text placement tendencies in mobile environments using video see-through displays. In: Proceedings of the IEEE international symposium on mixed and augmented reality (ISMAR), pp 281–282. <https://doi.org/10.1109/ismar.2013.6671805>
- Renkewitz H, Kinder V, Brandt M, Alexander T (2008) Optimal font size for head-mounted-displays in outdoor applications. In: Proceedings of the 12th international conference information visualization, pp 503–508. <https://doi.org/10.1109/iv.2008.73>
- Rosilius M, Wirsing B, von Eitzen I, Wilhelm M, Schmitt J, Engelmann B, Brautigam V (2021) Evaluation of Visual requirements and software-design for immersive visibility in industrial applications. In: Proceedings of IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct), pp 234–239. <https://doi.org/10.1109/ismar-adjunct54149.2021.00055>
- Rzayev R, Woźniak PW, Dingler T, Henze N (2018) Reading on smart glasses: The effect of text position, presentation type and walking. In: Proceedings of the 2018 CHI conference on human factors in computing systems, paper 45, pp 1–9. <https://doi.org/10.1145/3173574.3173619>
- Rzayev R, Ugriyenko P, Graf S, Schwind V, Henze N (2021) Reading in VR: The effect of text presentation type and location. In: Proceedings of the 2021 CHI conference on human factors in computing systems, pp 1–10. <https://doi.org/10.1145/3411764.3445606>
- Scharff LF, Ahumada AJ Jr, Hill AL (1999) Discriminability measures for predicting readability. In *Human Vision and Electronic Imaging IV* 3644:270–277. <https://doi.org/10.1117/12.348448>
- Schmalstieg D, Höllerer T (2016) *Augmented reality: Principles and practice*. Addison-Wesley, Boston
- Schopf J, Farace D (2010) Grey literature. In: Bates, M.J., Maack, M.N. (eds) *Encyclopedia of library and information sciences*, 3rd ed., CRC Press, London
- Shimizu Y, Ohnishi A, Terada T, Tsukamoto M (2021) Gaze-adaptive subtitles considering the balance among vertical/horizontal and depth of eye movement. In: Proceedings of IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct), pp 127–132. <https://doi.org/10.1109/ismar-adjunct54149.2021.00035>
- Sridharan SK, Hincapié-Ramos JD, Flatla DR, Irani P (2013) Color correction for optical see-through displays using display color profiles. In: Proceedings of the 19th ACM symposium on virtual reality software and technology, pp 231–240. <https://doi.org/10.1145/2503713.2503716>
- Tanaka K., Kishino Y., Miyamae M., Terada T., Nishio S. (2008) An Information layout method for an optical see-through head mounted display focusing on the viewability. In 7th IEEE/ACM international symposium on mixed and augmented reality, pp 139–142. <https://doi.org/10.1109/ISMAR.2008.4637340>
- Tsunajima T, Nishiuchi N (2020) Visual environment design of VR space for sequential reading in web browsing. In: Stephanidis, C., Antona, M. (eds) *HCI international 2020 - posters*. HCII 2020. Communications in computer and information science, vol 1225. https://doi.org/10.1007/978-3-030-50729-9_18
- Wang Z, Gao P, Ma L, Zhang W (2020) Effects of font size, line spacing, and font style on legibility of chinese characters on consumer-based virtual reality displays. In: Stephanidis, C., Antona, M., Ntoa, S. (eds) *HCI international 2020 – Late breaking posters*. HCII 2020. Communications in computer and information science, vol 1293. https://doi.org/10.1007/978-3-030-60700-5_59
- Wei C, Yu D, Dingler T (2020) Reading on 3D surfaces in virtual environments. In: Proceedings of the IEEE conference on virtual reality and 3D user interfaces (VR), pp 721–728. <https://doi.org/10.1109/vr46266.2020.00095>
- Woodward J, Smith J, Wang I, Cuenca S, Ruiz J (2020) Examining the presentation of information in augmented reality headsets for situational awareness. In: Proceedings of the international conference on advanced visual interfaces, pp 1–5. <https://doi.org/10.1145/3399715.3399846>
- Yuan J, Mansouri B, Pettey JH, Ahmed SF, Khaderi SK (2018) The visual effects associated with head-mounted displays. *Int J Ophthalmol Clin Res* 5(2):085. <https://doi.org/10.23937/2378-346x/1410085>
- Zhao Y, Hu M, Hashash S, Azenkot S (2017) Understanding low vision people's visual perception on commercial augmented reality glasses. In: Proceedings of the 2017 CHI conference on human factors in computing systems, pp 4170–4181. <https://doi.org/10.1145/3025453.3025949>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.