How Effective are Uni- and Multivariate Typographic Encodings? Studying the Usage of Font Weight, Oblique Angle, and Spacing

A. Bäuerle[1], R.Brath[2] , and M. El-Assady[3]

[1] University of Konstanz, Germany [2] Uncharted Software Inc., Canada [3] ETH Zurich, AI Center , Switzerland

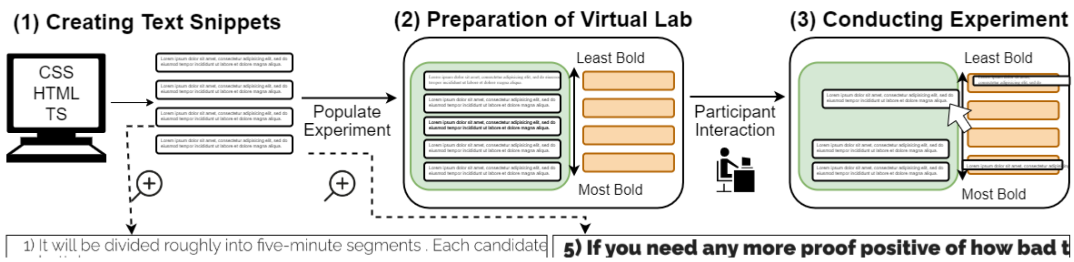


Figure 1: *Experiment workflow. After being annotated with typographic attributes, text snippets were exported as HD pictures. Participants helped us to evaluate the perceptual granularity of typographic attributes, by classifying the snippets, using a online collaboration platform.*

Abstract

Text is one of the most commonly used ways to transmit information. It is widely used in various visualizations and determines our understanding of the presented content. The information density of text can be enhanced by visualizing data in typographic attributes, such as font weight, letter spacing, or oblique angle. To increase information density the furthest, without the visualization losing performance or effectiveness, the perceivable granularity of the typographic attributes needs to be known. In an empirical experiment, the number of distinguishable levels in typographic attributes and the effects of changing the associated font size or facilitating multivariate encoding are assessed. Findings facilitate designing information-dense typographic visualizations without decreasing their performance or effectiveness.

1. Introduction

Data visualization is an interdisciplinary field that deals with transforming data into visual representations. An essential part of many of these visualizations is text. While text has mainly been separated out of visualizations [Bra20], text is capable of encoding data beyond the literal text. Utilizing these capabilities can yield more effective, more easily readable [Bei12, Bei09], and more information-dense visualizations.

Text can encode non-literal data by altering attributes associated with the typographic design space. These typographic attributes [AH06] include, amongst others:

- **Font Weight** (FW), refers to the stroke's thickness of letters.
- *Oblique Angle* (OA) refers to sloping text without modifying the letterforms.
- S p a c i n g (SP) modifies the horizontal gap between letters.

An appropriate visual encoding must be chosen to facilitate an advantageous data encoding into typographic attributes. This requires two key factors to be considered [SI11]: The natural order of a typographic attribute and the attribute's perceptual granularity. Perceptual granularity refers to how many distinct values of an attribute the user can meaningfully distinguish.

Addressing these requirements, we designed a study to evaluate the perceivable granularity of the typographic attributes mentioned above. Further, since multiple factors influence the number of distinguishable levels in a typographic attribute, we also investigated the influence of font size and multivariate encoding.

Typography, as a visualization technique, has been mostly neglected in the past decades, and thus knowledge about associated typographic attributes is limited. Our research is thus motivated by the pressing need to increase knowledge about text-associated visualization techniques, yielding more effective visualizations for text analysis. Concrete examples range from improving cartography on mobile devices [DRC16] to improving science notations, from creating more versatile code editors to more sophisticated graphical design and advertising.

Contributions – This paper makes the following contributions in the domain of visualizing data with text: (1) studying the perceptual granularity of the typographic attributes' FW, OA, and SP, evaluating existing theories on the attributes' perceptual granularity experimentally. In addition, we examine the (2) influence of font size on the perceptual granularity of typographic attributes. The impact of (3) combining typographic attributes is investigated as well. Further, we provide best practices for utilizing typographic attributes for encoding data, refining and complementing existing guides [SOK*15].

DOI: 10.2312/evs.20221085

2. Background, Related Work, and Problem Characterization

R. Brath [BB16] systematically explores and expands the design space of text visualizations. He provides a ranked overview of font-specific visual attributes, giving rough, theoretical estimates of the attributes' perceptual granularity. Our paper backs up or disproves these estimates with experimental evaluations. We extend his work further by considering the influence of font size on the perceptual granularity of typographic attributes.

Furthermore, C. Ware [War04] discusses the separability of visual variables and indicates combinations that are more easily perceptually separable or perceived as together (i.e. integral). In our paper, we examine the performance of typographic attributes used in multivariate visualizations, which allows us to evaluate the separability of the typographic attributes.

A challenge in visualization is to find the proper encoding for the data it represents; errors in choosing proper encodings are the most common [FJ10]. Steele and Iliinsky [SI11] state that a key factor for selecting a suitable visual encoding for an attribute is the number of distinguishable levels. This work is thus designed to examine the perceivable granularities of typographic attributes deeper. We provide proper usage recommendations for three typographic attributes at specific sizes and in multivariate usage scenarios.

Concepts and Terminology – In the scope of this paper, we define lower and upper bounds of typographic attributes as follows. For FW, we oriented ourselves on the CSS Fonts Level 4 specification, stating that the attribute FW ranges from 1 to 1000 [Big19]. There are, however, no proper standardized limitations for oblique angles. Modern fonts typically use slopes of up to 20 degrees. After investigating fonts in historical documents, which may be much steeper in slope, we decided the range to be -45 to 45 degrees. For spacing, the lower bound evaluates to 0 mm. Smaller values lead to overlapping characters, compromising the font's legibility. For readability reasons, we set the upper bound to be 10 mm.

Research Question and Hypotheses – In this paper, we strive to answer the following research question: "*What is the perceptual granularity of the typographic attributes FW, OA, and SP, and how does increasing the font size and the utilization of multivariate encoding affects the number of distinguishable levels?*"

In particular, we aim to investigate the following hypotheses:

(H1) **Influence of font size:** There is a significant positive relationship between font size and the perceptual granularity of FW, OA, and SP [HHMB12].

(H2) **Perceptual granularity:** According to the suitability of typographic attributes [Bra20], FW should have a higher perceptual granularity than OA. Based on the suitability for encoding data of the visual channels used by typographic encoding [SI11, Bra15], SP should have the highest number of distinguishable values.

(H3) **Multivariate typographic visualizations:** The effectiveness of typographic encoding deteriorates when the number of encoded values increases [Bri04]. Accordingly, the perceptual granularity of typographic attributes should decrease when used in multivariate visualizations.

3. Experimental Evaluation: User Study Setup

We conducted a quantitative within-subject user study to evaluate the problem stated above. Since boundaries for minimum and maximum values of typographic attributes' are difficult to determine, we designed the study to investigate the minimal distance between the attribute levels that can be meaningfully distinguished, called perceptual granularity distance, short PGD. After describing the study setup, we discuss its results and corresponding insights.

Methodology – Due to COVID-19-related restrictions, the study has been conducted online using the visual collaboration platform MIRO [Khu11]. MIRO was chosen because it facilitates the embedding of HD pictures and a collaborative real-time modification in a visual workspace. Since MIRO can sort trial templates, and due to the facilitator and the participant being in direct contact using ZOOM with screen share, proper remote guidance of the participant through the experiment was possible.

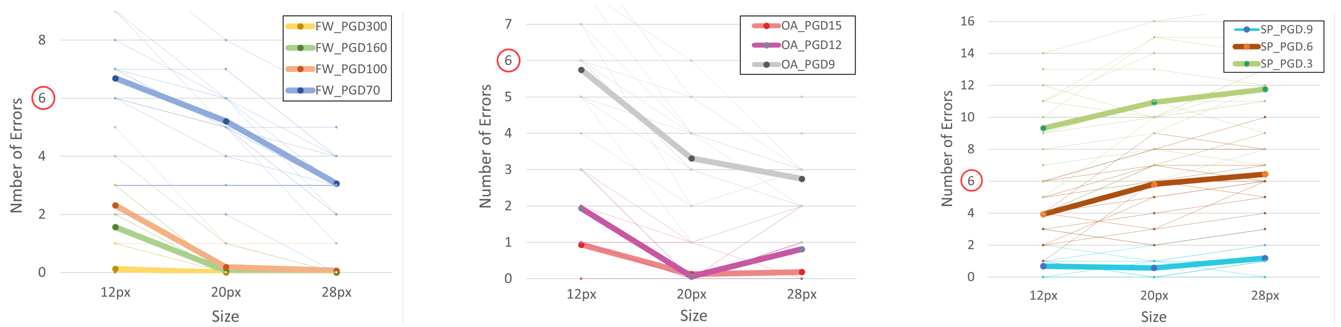
We conducted 16 three and a half hour sessions, in which the participants helped us evaluate the perceptual granularity distance of typographic attributes. Each session was divided into six parts. In the first 15 minutes, after welcoming the participant, we conducted a training. After ensuring the participant understood the basic MIRO controls and typographic attributes, we conducted four 40-minute sessions. The duration assigned to the sessions allowed for appropriately long breaks, compensating for fatigue effects. These sessions included 36 trials in which the participant classified and ordered levels of granularity of typographically modified text snippets. The difference distance between levels has been varied, see Table 1:

TA	Abbr. Prefix	CSS-Attr.	Level dist.
FW	FW_PGD	font-weight	300,160,100,70
OA	OA_PGD	-moz-transform	15°, 11°, 9°
SP	SP_PGD	letter-spacing (mm)	0.9, 0.6, 0.3

Table 1: ***T****ypographic* ***A****ttributes examined, their* ***abbr****eviations, the css-attributes modified and the difference between levels used.*

Each case above has been conducted for the font sizes 12, 20, and 28 px (2.5, 4.125, and 5.75 mm). The font sizes were selected based on their popularity, with 12 px being the most commonly used font size for text bodies [Eil22] and 20 px being the most commonly used font size for headlines [Mar09]. By extrapolation, 28 px was obtained, marking the upper bound limit for often used font sizes [Mar09]. The variable font used is *Raleway*, designed by Matt McInerney [MIF16]. The multivariate part of our experiment was conducted with a font size of 20 px. Each possible bivariate combination of FW_PGD300, OA_PGD15, and AP_PGD.9 was combined and existed twice. The resulting utterances were exported as HD pictures and embedded into MIRO so that zooming in was only possible to a limited degree. This assured that, with the maximum zoom applied, the above-stated sizes were not exceeded. The classification process was facilitated by using templates similar to Figure 1. For each trial, we measured the time taken, the number of classification errors, and the score of a modified version of an unweighted NASA TLX [CCPE09]. After the experiment, the classification errors were measured by counting the wrongly classified text snippets, which were placed in an incorrect box. The NASA TLX was chosen for its ease of use, reliable sensitivity, and suitability for measuring a subjective workload assessment.

Datasets and Controls – To ensure the validity and reliability of our online-study, we varied the content and content length of the text snippets. To hamper the classification of text snippets based on features we do not want to examine and to reflect the majority of use cases regarding linguistic features, the content length was as



(a) Font Weight **(b)** Oblique Angle **(c)** Spacing

Figure 2: *Study Results, depicting the number of size-dependent classification errors. Thin lines correspond to a single participant and PGD. Thick lines display the population average. Note, that axis scales are nonidentical; thus, the error rate of 6 is highlighted to ease comparability.*

long as small utterances. Since the training session included the text snippets, the participant was familiar with their content, limiting it as a confounding factor. We further randomized the granularity level applied to the text snippets and the order in which text blocks were added to the text block container (green box in Figure 1), short TBC. This is only restricted such that each level needs to be present at least once. The sequence in which the different cases have been evaluated has also been randomized to compensate for learning effects. Screen size and resolution were restricted to 24 inches (0.61 m) and at least 1920×1080 pixels. A good working mouse was a prerequisite.

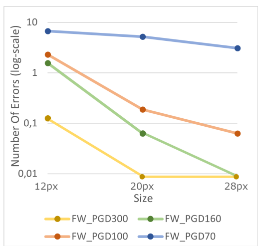
Participants – The 16 participants in our study, out of which six were female, were aged above 20 and below 28 years (∅24.7 years) and had to be fluent in English. Visual impairments had to be corrected back to normal. None of them had experience with MIRO or the perceptual granularity of typographic attributes. All participants work with a PC regularly (∅ 3.1 hours per day).

Tasks – Based on our problem description, the task was to classify and sort 22 text blocks formulated in English and modified according to the independent variables of font size, granularity, and combination of typographic attributes, from the TBC section into the Sorted Classification Area, short SCA (orange boxes in Figure 1). This ranking task allowed us, as opposed to other methods (e.g., staircase-method [Cor62]), to present the participant with stimuli in an unordered way, reducing distress when judgments become complex. For the multivariate classification, the contained text snippets were non-redundantly and independently annotated with two typographic attributes. To support the result comparability, each bivariate combination existed twice as an individual template, each of which had to be sorted based on only one of the attributes.

4. Experimental Evaluation: User Study Results

Using the boundaries from section 2 and taking the effective PGDs into account, the corresponding perceptual granularity can be determined. Conducting the experiment yielded the following results.

Font Weight – As depicted in Figure 2a, increasing font size helps in reducing the classification error rate for different perceptual granularity levels of FW. For every perceptual granularity investigated, increasing the font size entails a decrease in the error rate, faster task completion time, and a lower perceived workload. Based on the Weber-Fechner laws [Hec24], both relating to human perception, the relationship between stimulus and perception is logarithmic. Applied to our FW study, if the distance between levels of the stimulus is reduced, error rates should rise in an exponential manner. The side-figure maps the error rate onto a logarithmic scale, allowing this effect to be easily observed. According to our study results, at a 28 px font size, a distance of 100 between the attribute levels can be effectively distinguished (see Figure 2a). Considering the lower and upper bound of FW, ten levels can be used at a 28 px font size. More than ten levels are not a valid option due to the decrease in effectiveness and efficiency and increased workload. Consequently, a difference of 100 in the CSS-property "font-weight" is the limit. However, the common usage of only four equally spaced levels is justified. Due to its shallow error rate, especially for small font sizes, using four FWs is a broadly applicable perceptual granularity. However, more granular depictions are a more complex but beneficial option.



Oblique Angle – The perceptual granularity of OA behaves differently. While increasing the font size improved the efficiency and effectiveness of FW, increasing the font size for OA improves its performance only to a specific granularity-dependent size before it deteriorates or does not improve further. After reaching that font size, no significant increase in performance can be observed. The smaller the angle difference, the bigger the font size must be to reach the point where no significant increase is observed.

As depicted in Figure 2b, angle differences around nine degrees (OA_PGD9) are very error-prone for font sizes up to 28 px, rendering its usage ineffective when ordered with 12 degrees. Differences of 12 and 15 degrees reach their plateau at a font size of around 20 px. Amplifying font size further does not increase performance.

Spacing – Opposed to the findings discussed above, the typographic attribute SP behaves, regarding **H1**, contrary to FW. Increasing the font size worsens the extent to which the participants can distinguish spacing versions (see Figure 2c). Small font sizes are better suited to distinguish more levels of perceptual granularity. Differences below 0.9 mm between the spacing levels are not effective.

Multivariate Encoding – Table 2 depicts the relative change of performance regarding the identification of granularity classes of an attribute in a use case of multivariate encoding towards a use case of univariate encoding. Advantageous changes, such as a reduction in the task completion time, perceived workload, or average error frequency, are emphasized using blue color variations, and disadvantageous changes are emphasized using red color variations. The saturation level corresponds to the extent of the relative performance change:

	Time	TLX	Error	Time	TLX	Error
	FW+OA			F W + S P		
FW	29.47	15.87	0	70.14	73.85	0
	OA+FW			*O A + S P*		
OA	14.37	79.87	478.4	10.17	105.91	3580
	S P + F W			*S P + O A*		
SP	17.29	23.68	304.6	27.81	22.18	554.86

Table 2: *Effectiveness change (in %), comparing effectiveness of typographic attributes in bivariate versus univariate encodings.*

Multivariate encoding increases the information density of typographic visualizations. However, the excessive coverage of red and the absence of the blue color range in Table 2 indicates that using multiple typographic attributes in the same spatial dimension does dramatically worsen the effectiveness and efficiency of multivariate typographic visualizations. Apart from small increases in perceived workload and time taken to complete the task, FW is little affected, at least for low granularities. OA and SP perform worse when used in multivariate typographic encoding. The effectiveness of both OA and SP are way worse when combined with SP, respectively OA, as opposed to combining them with FW.

5. Discussion and User Study Interpretation

Based on the observations made, after interpreting the results and explaining occurring phenomena, we supplement and enhance existing guides for the usage of typographic attributes.

Lessons Learned – As results have shown, the perceptual granularity heavily depends on the font size used. Opposed to **H1**, increasing font size has, depending on the typographic attribute used, a varying influence on the perception of perceptual granularity levels.

The negative influence of increasing the font size on SP can be explained by our usage of absolute distances, whose relative spacing width decreases compared to in-word lengths, thus decreasing in significance. The initially positive influence of font size increases on OA is explained by increasing the visibility of associated angles. The following decrease in performance, when increasing size further, can be explained by influencing factors such as increased arm lengths, perceived angles, or the reduced number of letters visible in parafoveal vision. According to Wenderroth et al. [WJ84], the angle's arm lengths distort the angle perception, compounding the effectiveness of increasing font size. Also, at larger sizes, there are fewer letters in the area of parafoveal vision. Thus, there may be more scanning required to identify letters with long stems (e.g. b,q,k), which are more significant for judging angle, than rounded letters with no stems (e.g. e,c), and with increased scanning, the time and errors may increase. The advantageous effects on FW of increasing font size result from the increasing distances between single levels of perceptual granularity, both in the widths of the letter strokes and the intensity of utterances.

Results show that classification tasks of a single typographic attribute in a multivariate visualization generally yield less effective encoding compared to univariate visualization, confirming our hypothesis **H3**. Opposed to FW, both OA and SP are perceived holistically [Gar74] and are thus hard to decode from combined visual variables. However, our participants could make separate and concrete judgments about FW. The attribute can still be perceptually distinguished in multivariate visualizations. We thus assign FW to the separable and both SP and OA to integral dimension [War04].

The number of perceptual granularity levels used needs to be adapted to the font size used to create a compelling visualization.

Best Practices – Extending the ranked overview of font-specific visual attributes [Bra20, Bra15] and the guide of Steele and Iliinsky et al. [SI11] towards typographic attributes, the following recommendations towards the perceptual granularity can be made:

For FW, for font sizes smaller than 20 px, the cardinality of the perceptual granularity evaluates to four. For 20 px or higher, up to ten levels can be used. Regarding OA, we suggest not using more than seven levels and suggest checking readability of steeper slopes for the target task. Font size should be chosen big enough, such that readability and a clear visualization are not compromised, but not too big, such that performance suffers. A difference below 0.9 mm between the distinct levels should not be used for SP. Font sizes used for SP, at least for absolute values, should be chosen as small as possible without interfering with readability. Considering our boundaries, discussed in section 2, we come up with up to 11 levels. Our results confirm our hypothesis **H2**, and further specify the work by Brath [BB16], who derived two to nine levels for FW and two to five levels for OA. We extended the work with estimates for SP.

Regarding its effectiveness, FW is a strong perceptual [BB14] cue and also highly separable. Thus, if one needs only one typographic attribute, encode the variables using FW. Further assignments are an open question, with the decision depending on multiple factors.

Limitations – The cardinality of a typographic attribute's perceptual granularity heavily depends on the lower and upper bounds of the respective attribute. These bounds were determined based on historical usages, CSS-attribute boundaries, and legibility issues, missing experimental validation. Further, the presented results on the perceptual granularity of typographic attributes heavily depend on the controlled conditions of our online experiment. A variety of conditions,such as display types [SSZH05], screen resolution [Wil03], different pixel rendering techniques [RE18, FE15, MS12, Big19], different text constraints [BGN08, BBK*18] based on close- or distant-reading views, and font type used, may influence the corresponding cardinality for real-world use cases drastically. Additionally, we only evaluated a subset of the entirety of typographic attributes.

Future Work – To increase the applicability of our results, boundaries of typographic attributes should be assessed experimentally, such that the readability [Bur49] and the legibility at-a-glance [SDCR20] of text are not impaired. Assessing legibility issues will allow us to determine the point of "critical letter-spacing" [Big19] of different font types as well as extend the current research around the legibility of oblique angles, which is often limited to italic [Big19]. We further consider expanding the study onto other typographic variables, font sizes, and font types. Since font types affect low vision reading [Leg06], we plan to assess the effects of different font types and low vision on the perceived perceptual granularity. Considering the small number of participants, this study will be re-conducted on a larger scale.

6. Conclusion

Finding the proper encoding for data is a very error-prone steps in designing data visualizations. Facilitating the correct choice, we conducted a study examining the perceptual granularity of typographic attributes and the corresponding effects of font size and multivariate data encoding. As a result, we confirm prior theoretical findings, discover new properties of typographic attributes regarding font size, and provide best practices for using typographic attributes.

References

[AH06] AMBROSE G., HARRIS P.: The Fundamentals of Typography. Ava Publishing, pp. 80–101. 1

[BB14] BRATH R., BANISSI E.: The Design Space of Typeface. In *IEEE Conf. on Information Visualisation VIS* (2014), p. 1. 4

[BB16] BRATH R., BANISSI E.: Using Typography to expand the Design Space of Data Visualization. *She Ji: The J. of Design, Economics, and Innovation 2*, 1 (2016), 59–87. 2, 4

[BBK*18] BEHRISCH M., BLUMENSCHEIN M., KIM N. W., SHAO L., EL-ASSADY M., FUCHS J., SEEBACHER D., DIEHL A., BRANDES U., PFISTER H., ET AL.: Quality Metrics for Information Visualization. In *Computer Graphics Forum* (2018), vol. 37, Wiley Online Library, pp. 625–662. 4

[Bei09] BEIER S.: Typeface Legibility: Towards Defining Familiarity. Royal College of Art (United Kingdom), pp. 22–27. 1

[Bei12] BEIER S.: Reading Letters: Designing for Legibility. Bis Publishers, pp. 31–36. 1

[BGN08] BATEMAN S., GUTWIN C., NACENTA M.: Seeing Things in the Clouds: The Effect of visual Features on Tag Cloud Selections. In *Proc. ACM Conf. on Hypertext and Hypermedia* (2008), pp. 193–202. 4

[Big19] BIGELOW C.: Typeface Features and Legibility Research. *Vision research 165* (2019), 162–172. 2, 4

[Bra15] BRATH R.: Meta Ranking of Visual Attributes in Data Visualization, 2015. URL: https://richardbrath.wordpress.com/2015/10/05/meta-ranking-of-visual-attributes-in-data-visualization/. 2, 4

[Bra20] BRATH R.: Visualizing with Text. CRC Press, pp. 43–66. 1, 2, 4

[Bri04] BRINGHURST R.: The Elements of typographic Style. Hartley & Marks Vancouver, pp. 95–103. 2

[Bur49] BURTT H. E.: Typography and Readability. *Elementary English 26*, 4 (1949), 212–221. 4

[CCPE09] CAO A., CHINTAMANI K. K., PANDYA A. K., ELLIS R. D.: NASA TLX: Software for assessing subjective mental Workload. *Behavior Research Methods 41*, 1 (2009), 113–117. 2

[Cor62] CORNSWEET T. N.: The Staircase-method in Psychophysics. *The American journal of psychology 75*, 3 (1962), 485–491. 3

[DRC16] DOBRES J., REIMER B., CHAHINE N.: The Effect of Font Weight and rendering System on glance-based Text Legibility. In *Proc. of the Int. Conf. on Automotive User Interfaces and Interactive Vehicular Applications* (2016), pp. 91–96. 1

[Eil22] EILERS C.: Best font for resume: Size, standard professional pairings, 2022. URL: https://zety.com/blog/best-fonts-for-resume. 2

[FE15] FIFIELD D., EGELMAN S.: Fingerprinting Web Users through Font Metrics. In *In. Conf. on Financial Cryptography and Data Security* (2015), Springer, pp. 107–124. 4

[FJ10] FORSELL C., JOHANSSON J.: An Heuristic Set for Evaluation in Information Visualization. In *Proc. of the Workshop on Advanced Visual Interfaces AVI* (2010), pp. 199–206. 2

[Gar74] GARNER W. R.: The Stimulus in Information Processing. In *Sensation and Measurement*. Springer, 1974, pp. 77–90. 4

[Hec24] HECHT S.: The visual discrimination of Intensity and the Weber-Fechner Law. *The J. of General Physiology 7*, 2 (1924), 235. 3

[HHMB12] HEATH M., HOLMES S. A., MULLA A., BINSTED G.: Grasping Time does not influence the early Adherence of Aperture shaping to Weber's Law. *Frontiers in Human Neuroscience 6* (2012), 332. 2

[Khu11] KHUSID A.: Miro - The leading visual Collaboration Platform, 2011. URL: http:/https://miro.com/about/. 2

[Leg06] LEGGE G. E.: Psychophysics of Reading in normal and low Vision. CRC Press, pp. 45–51. 4

[Mar09] MARTIN M.: Typographic Design Patterns and Best Practices, 2009. URL: https://www.smashingmagazine.com/2009/08/typographic-design-survey-best-practices-from-the-best-blogs/. 2

[MIF16] MCINERNEY M., IMPALLARI P., FUENZALIDA R.: Google Fonts - Raleway, 2016. URL: https://fonts.google.com/specimen/Raleway. 2

[MS12] MOWERY K., SHACHAM H.: Pixel perfect: Fingerprinting Canvas in HTML5. *Proc. of W2SP* (2012), 5–6. 4

[RE18] ROUGIER N. P., ESFAHBOD B.: Digital Typography Rendering. In *ACM SIGGRAPH Courses* (2018), pp. 1–3. 4

[SDCR20] SAWYER B. D., DOBRES J., CHAHINE N., REIMER B.: The great Typography Bake-off: Comparing Legibility at-a-glance. *Ergonomics 63*, 4 (2020), 391–398. 4

[SI11] STEELE J., ILIINSKY N.: Designing Data Visualizations. O'Reilly Media, Inc., pp. 25–29. 1, 2, 4

[SOK*15] STROBELT H., OELKE D., KWON B. C., SCHRECK T., PFISTER H.: Guidelines for effective Usage of Text Highlighting Techniques. *IEEE Trans. Visualization and Computer Graphics 22*, 1 (2015), 489–498. 1

[SSZH05] SHEEDY J. E., SUBBARAM M. V., ZIMMERMAN A. B., HAYES J. R.: Text Legibility and the Letter Superiority Effect. *Human factors 47*, 4 (2005), 797–815. 4

[War04] WARE C.: *Information Visualization: Perception for Design: Second Edition*. 2004, pp. 160–168. 2, 4

[Wil03] WILLIAMS G.: Font Creation with FontForge. *EuroTEX 2003 Proc., TUGboat 24*, 3 (2003), 531–544. 4

[WJ84] WENDEROTH P., JOHNSON M.: The Effects of Angle-Arm Length on Judgments of Angle Magnitude and Orientation Contrast. *Perception & Psychophysics 36*, 6 (1984), 538–544. 4